DUDLEY KNOX LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY CALIFORNIA 93943-5002

# NAVAL POSTGRADUATE SCHOOL
## Monterey, California

# THESIS

WATER MASS AND ACOUSTIC ANALYSIS
OF THE EAST GREENLAND CURRENT

by

James M. Clipson

September 1986

Thesis Co-Advisors: R.H. Bourke
L.C. Breaker

Approved for public release; distribution is unlimited

T230169

Unclassified
ECURITY CLASSIFICATION OF THIS PAGE

# REPORT DOCUMENTATION PAGE

| | |
|---|---|
| a REPORT SECURITY CLASSIFICATION: Unclassified | 1b RESTRICTIVE MARKINGS: |
| a SECURITY CLASSIFICATION AUTHORITY | 3 DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution is unlimited. |
| b DECLASSIFICATION / DOWNGRADING SCHEDULE | |
| PERFORMING ORGANIZATION REPORT NUMBER(S) | S MONITORING ORGANIZATION REPORT NUMBER(S) |
| a NAME OF PERFORMING ORGANIZATION: Naval Postgraduate School; 6b OFFICE SYMBOL (If applicable): 68 | 7a NAME OF MONITORING ORGANIZATION: Naval Postgraduate School |
| c. ADDRESS (City, State, and ZIP Code): Monterey, California 93943-5000 | 7b. ADDRESS (City, State, and ZIP Code): Monterey, California 93943-5000 |
| a NAME OF FUNDING / SPONSORING ORGANIZATION; 8b OFFICE SYMBOL (If applicable) | 9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER |
| : ADDRESS (City, State, and ZIP Code) | 10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO; PROJECT NO; TASK NO; WORK UNIT ACCESSION NO |

TITLE (Include Security Classification): WATER MASS AND ACOUSTIC ANALYSIS OF THE EAST GREENLAND CURRENT — UNCLASSIFIED

PERSONAL AUTHOR(S): Clipson, James M.

| a TYPE OF REPORT | 13b TIME COVERED | 14 DATE OF REPORT (Year, Month, Day) | 15 PAGE COUNT |
|---|---|---|---|
| Masters Thesis | FROM _____ TO _____ | 1986 September | 97 |

SUPPLEMENTARY NOTATION

| COSATI CODES | | | 18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number) |
|---|---|---|---|
| FIELD | GROUP | SUB-GROUP | Cluster Analysis Entity Attribute Separation Distance Acoustic Analysis Propagation Loss Ducting Ray Tracing |
| | | | |


ABSTRACT (Continue on reverse if necessary and identify by block number)

vo cluster analysis techniques, one heuristic and one iterative, are employed to investi-
ate the spatial coherence of the water masses of the East Greenland Current (EGC), in the
icinity of the East Greenland Polar Front (EGPF). Both techniques are shown to be gener-
lly reliable, although the iterative technique is more consistent with classical oceanog-
aphic analyses. The techniques are applied to data to explore the grouping behaviour of
ne water masses. They are shown to have applications in planning a sonobuoy pattern and
ssessing the validity of XBT data prior to an acoustic forecast.

coustical analysis shows that acoustic reciprocity does not hold for propagation in the
aters of the EGC. Ranges from shallow to deep water are far in excess of those from deep
o shallow water. Propagation across the EGPF is significantly different for normal and
blique cases. Propagation loss for oblique ranges is between 60 and 80% of perpendicular
anges, mostly due to different source sound speed profiles. Three acoustic models, FACT,
AYMODE and PE are compared and contrasted. PE is found to be the most consistent and


| | |
|---|---|
| DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS | 21 ABSTRACT SECURITY CLASSIFICATION: Unclassified |
| a NAME OF RESPONSIBLE INDIVIDUAL: H Bourke | 22b TELEPHONE (Include Area Code): 408 646 3270 — 22c OFFICE SYMBOL: 68Bf |

D FORM 1473, 84 MAR — 83 APR edition may be used until exhausted; All other editions are obsolete — SECURITY CLASSIFICATION OF THIS PAGE

19. Abstract (continued)

..... reliable, although both FACT and RAYMODE compare satisfactorily for propagation from shallow to deep water. However, for the reverse case, FACT overestimates ranges by a factor of two, whereas RAYMODE is exceedingly over optimistic in its forecast ranges.

Approved for public release; distribution is unlimited.

# Water Mass and Acoustic Analysis of the East Greenland Current

by

James Michael Clipson
Lieutenant-Commander Royal Navy
B.A., Open University, 1975
M.Sc., Teesside Polytechnic, 1978

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OCEANOGRAPHY

from the

NAVAL POSTGRADUATE SCHOOL
September 1986

## ABSTRACT

Two cluster analysis techniques, one heuristic and one iterative, are employed to investigate the spatial coherence of the water masses of the East Greenland Current (EGC), in the vicinity of the East Greenland Polar Front (EGPF). Both techniques are shown to be generally reliable, although the iterative technique is more consistent with classical oceanographic analyses. The techniques are applied to data to explore the grouping behaviour of the water masses. They are also shown to have applications to multiple and single variable data. The cluster technique is shown to have applications in planning a sonobuoy pattern and in assessing the validity of XBT data prior to an acoustic forecast.

Acoustical analysis shows that acoustic reciprocity does not hold for propagation in the waters of the EGC. Ranges from shallow to deep water are far in excess of those from deep to shallow water. Propagation across the EGPF is significantly different for normal and oblique cases. Propagation loss for oblique ranges is between 60 and 80% of perpendicular ranges, mostly due to different source sound speed profiles. Three acoustic models, FACT, RAYMODE and PE are compared and contrasted. PE is found to be the most consistent and reliable, although both FACT and RAYMODE compare satisfactorily for propagation from shallow to deep water. However, for the reverse case, FACT overestimates ranges by a factor of two, whereas RAYMODE is exceedingly over optimistic in its forecast ranges.

TABLE OF CONTENTS


I. INTRODUCTION ..... 12
A. BACKGROUND ..... 12
B. PHYSICAL OCEANOGRAPHY ..... 12
C. SPATIAL COHERENCE ..... 14
D. CLUSTER ANALYSIS ..... 15
E. ACOUSTICAL ANALYSIS ..... 16

II. CLUSTER ANALYSIS ..... 19
A. DETAILS OF CLUSTER ANALYSIS ..... 19
B. DISTANCE ..... 20
C. THE DATA SET ..... 20
D. HEURISTIC TECHNIQUE ..... 21
E. A SECOND SIMULATION ..... 23
F. K-MEANS ..... 28
G. SUMMARY ..... 33

III. APPLICATIONS OF CLUSTER ANALYSIS ..... 34
A. INTRODUCTION ..... 34
B. T-S ANALYSIS OF THE EAST GREENLAND CURRENT ..... 34
C. CLUSTER ANALYSIS ..... 38
D. ENTIRE DATA SET ..... 39
1. The mixed layer ..... 39
2. 150 m ..... 39
E. STATIONS IN THE VICINITY OF THE FRONT ..... 46
1. 10-20 m ..... 46
2. 500 m ..... 46
F. FRONTAL TRANSECTS ..... 51
G. WARM STATIONS ..... 59
H. SUMMARY ..... 63

IV. ACOUSTICAL ANALYSIS ..... 68
A. INTRODUCTION ..... 68
B. THE ACOUSTIC MODELS ..... 68
1. The PE Model ..... 68
2. The RAYMODE Model ..... 69
3. The FACT 9H Model ..... 70
C. ANALYSIS ..... 70
1. Source - Receiver Dispositions ..... 72
D. TRANSMISSION LOSS ..... 77
1. From Deep Water (looking shoreward) ..... 77
2. From Shallow Water (looking east) ..... 79
E. DISCUSSION ..... 81
1. Model Differences ..... 81
2. Normal and Oblique Transects ..... 83
3. Acoustic Reciprocity ..... 84
4. Conclusions ..... 89

V. SUMMARY AND CONCLUSIONS ..... 90

LIST OF REFERENCES ..... 92

INITIAL DISTRIBUTION LIST ..... 94

LIST OF TABLES


I. SIX TEMPERATURE - SALINITY PAIRS SIMULATED DATA ..... 21

II. INPUT TO THE LEADER ALGORITHM ..... 22

III. OUTPUT OF THE LEADER ALGORITHM ..... 26

IV. INPUT FOR THE K-MEANS TECHNIQUE ..... 31

V. TWO-CLUSTER OUTPUT FOR K-MEANS TECHNIQUE ..... 31

VI. THREE-CLUSTER OUTPUT FOR K-MEANS TECHNIQUE ..... 32

VII. LOCATION AND MEAN T-S VALUES AT 10-20 M FOR STATIONS 196-201 ..... 56

VIII. T-S VALUES FOR STATIONS 196-201 (150 M) ..... 59

IX. PREDICTED RANGES (FACT) ..... 78

X. PREDICTED RANGES (RAYMODE) ..... 78

XI. PREDICTED RANGES (PE) ..... 79

XII. PREDICTED RANGES (PE) ..... 79

XIII. PREDICTED RANGES (PE) ..... 80

XIV. PREDICTED RANGES (FACT) ..... 80

XV. PREDICTED RANGES (RAYMODE) ..... 80

XVI. PREDICTED RANGES (PE) ..... 81

XVII. PREDICTED RANGES (PE) ..... 82

XVIII. PREDICTED RANGES (PE) ..... 82

## LIST OF FIGURES


1.1 A map showing the general bathymetry and circulation in the Greenland Sea (Paquette et al., 1985, p. 4867) ..... 13

1.2 A plot of (a) compact clusters, and (b) elongated clusters (Chatfield and Collins, 1980, p.217) ..... 17

2.1 A plot of separation distance versus the minimum number of clusters for the simulated data ..... 24

2.2 A plot of separation distance versus the minimum number of clusters for the three sets of random data ..... 25

2.3 A best fit curve for the random data ..... 27

2.4 An illustration of the area of interest when considering random data ..... 29

2.5 Steps in the optimum clustering of 10 data points (Spath, 1980, p. 31) ..... 30

3.1 A map of the oceanographic stations forming the data set. Two frontal transects used in the cluster analysis are indicated. The EGPF is also shown ..... 35

3.2 A T-S plot of an AIW station (201) to the east of the EGPF, a PW station (247) on the shelf and a further PW station (210) in the frontal region ..... 36

3.3 A T-S plot of two shelf stations. Station 225 is situated on the western part of the shelf. Station 310 is situated to the east of the shelf ..... 37

3.4 Results of a two-cluster search in the mixed layer using the heuristic and iterative techniques ..... 40

3.5 Results of a three-cluster search in the mixed layer using the heuristic technique ..... 41

3.6 Results of a three-cluster search in the mixed layer using the iterative technique ..... 42

3.7 Results of two-cluster search at 150 m using the heuristic and iterative techniques ..... 43

3.8 Results of a three-cluster search at 150 m using the heuristic technique ..... 44

3.9 Results of a three-cluster search at 150 m using the iterative technique ..... 45

3.10 Results of a two-cluster search in the mixed layer for Frontal Stations using the heuristic and iterative techniques ..... 47

3.11 Results of a three-cluster search in the mixed layer for Frontal Stations using the heuristic technique ..... 48

3.12 Results of a three-cluster search in the mixed layer for Frontal Stations using the iterative technique ..... 49

3.13 Results of a two-cluster search at 500 m for Frontal Stations using the heuristic and iterative techniques ..... 50

3.14 Results of a three-cluster search at 500 m for Frontal Stations using the heuristic technique ..... 52

3.15 Results of a three-cluster search at 500 m for Frontal Stations using the iterative technique ..... 53

3.16 Results of a two-cluster search in the mixed layer for Stations 196-201 using the heuristic and iterative techniques ..... 54

3.17 Results of a three-cluster search in the mixed layer for Stations 196-201 using the heuristic and iterative techniques ..... 55

3.18 Results of a two-cluster search , using three attributes for Stations 196-201 ..... 57

3.19 Results of a three-cluster search , using three attributes for Stations 196-201 ..... 58

3.20 Results of a two-cluster search at 150 m for Stations 196-201 using the heuristic and iterative techniques ..... 60

3.21 Results of a three-cluster search at 150 m for Stations 196-201 using the heuristic and iterative techniques ..... 61

3.22 A plot of temperature, salinity, sigma-t and sound speed for a typical AIW station (Bourke and Paquette, 1985) ..... 62

3.23 Results of a two-cluster search at a sigma-t value of 27.8 for AIW stations using the heuristic and iterative techniques ..... 64

3.24 Results of a three-cluster search at a sigma-t value of 27.8 for AIW stations using the heuristic technique ..... 65

3.25 Results of a three-cluster search at a sigma-t value of 27.8 for AIW stations using the iterative technique ..... 66

4.1 A map to show the position of the two frontal transects used for the acoustical analysis. The EGPF is also shown ..... 69

4.2 A property plot of Station 270, in deep AIW to the east of the EGPF (Bourke and Paquette, 1985) ..... 72

4.3 A property plot of Station 277, in shallow PW on the continental shelf (Bourke and Paquette, 1985) ..... 73

4.4 A property plot of Station 283, in deep AIW to the east of the EGPF (Bourke and Paquette, 1985) ..... 74

4.5 A sound speed plot of Stations 270 - 277, a perpendicular transect. Sound speed at the bottom of the water column is indicated ..... 75

4.6 A sound speed plot of Stations 277 - 283, an oblique transect. Sound speed at the bottom of the water column is indicated ..... 76

4.7 A PL curve for PE at 50 Hz using multiple sound speed profiles from a perpendicular transect (east to west). The source is at 60 m, the receiver is at 60 m ..... 85

4.8 A PL curve for PE at 50 Hz using multiple sound speed profiles from a perpendicular transect (west to east). The source is at 60 m, the receiver is at 60 m ..... 86

4.9 A PL curve for PE at 50 Hz using multiple sound speed profiles from a perpendicular transect (west to east), assuming a constant water depth. The source is at 60 m, the receiver is at 60 m . . . . . 87

4.10 A PL curve for PE at 50 Hz using a single sound speed profile, assuming a sloping ocean bottom. The source is at 60 m, the receiver is at 60 m . . . . . 88

## ACKNOWLEDGEMENTS

I would like to thank my two advisors for their encouragement, advice and concern throughout the production of this study. Dr. Robert Bourke and Dr. Laurence Breaker have been excellent mentors at all times. My grateful thanks go to Dr. Robert Paquette, Ms. Arlene Bird and Ms. Donna Burych for their invaluable help in programming and also to Mr. Larry Frazier for his assistance in dealing with thesis software. My fellow students have been a great help in providing aid and encouragement, particular gratitude is due to; Skip Beale, Gib Kerr, Phil Renaud, John Rendine, Steve Summers and Alan Weigel.

I would also like to offer my thanks to the Director of Naval Oceanography and Meteorology (London) for giving me this unique opportunity to study at the US Naval Postgraduate School in Monterey.

# I. INTRODUCTION

## A. BACKGROUND

The East Greenland Sea area has attracted considerable oceanographic interest in recent years. One reason for this is the area's strategic importance for NATO; another is the great interest now being shown in Arctic waters by Naval planners and strategists.

The data obtained from the MIZLANT 84 cruise (Bourke and Paquette, 1985) and previous similar cruises have provided the basis for a physical oceanographic analysis of the waters overlying the East Greenland continental shelf and slope (Tunnicliffe, 1985). It has provided an opportunity for investigation into acoustic propagation across the ocean front found at the ice edge, the East Greenland Polar Front (EGPF) (Sleichter, 1984). This study draws on the data obtained during the MIZLANT 84 cruise. The EGPF has been identified by previous oceanographic analyses and it is the purpose of this research to analyse the frontal region and adjacent water masses by statistical methods and to investigate the possible uses of these methods in other oceanographic regions. The statistical method used is cluster analysis. Cluster analysis is a broad term given to techniques that group entities into homogeneous subgroups on the basis of their similarities (Lorr, 1983). In addition, the study will conduct an acoustic analysis in the frontal region, using three acoustic models that are currently in operation or are at an advanced research stage. Before the statistical analysis techniques are discussed, the oceanographic background will be briefly described.

## B. PHYSICAL OCEANOGRAPHY

The water masses of the East Greenland Current have been identified by Aagaard and Coachman (1968a and 1968b) and these definitions are adopted here. The following description of the physical oceanography of the region is taken from Tunnicliffe (1985).

The circulation pattern in the Greenland Sea is shown in Figure 1.1 (Paquette et al, 1985). This figure shows that the surface circulation is a large cyclonic gyre bounded by the Jan Mayen Current to the south and the Norwegian and West Spitsbergen Currents to the east. In the north the West Spitsbergen Current (WSC)

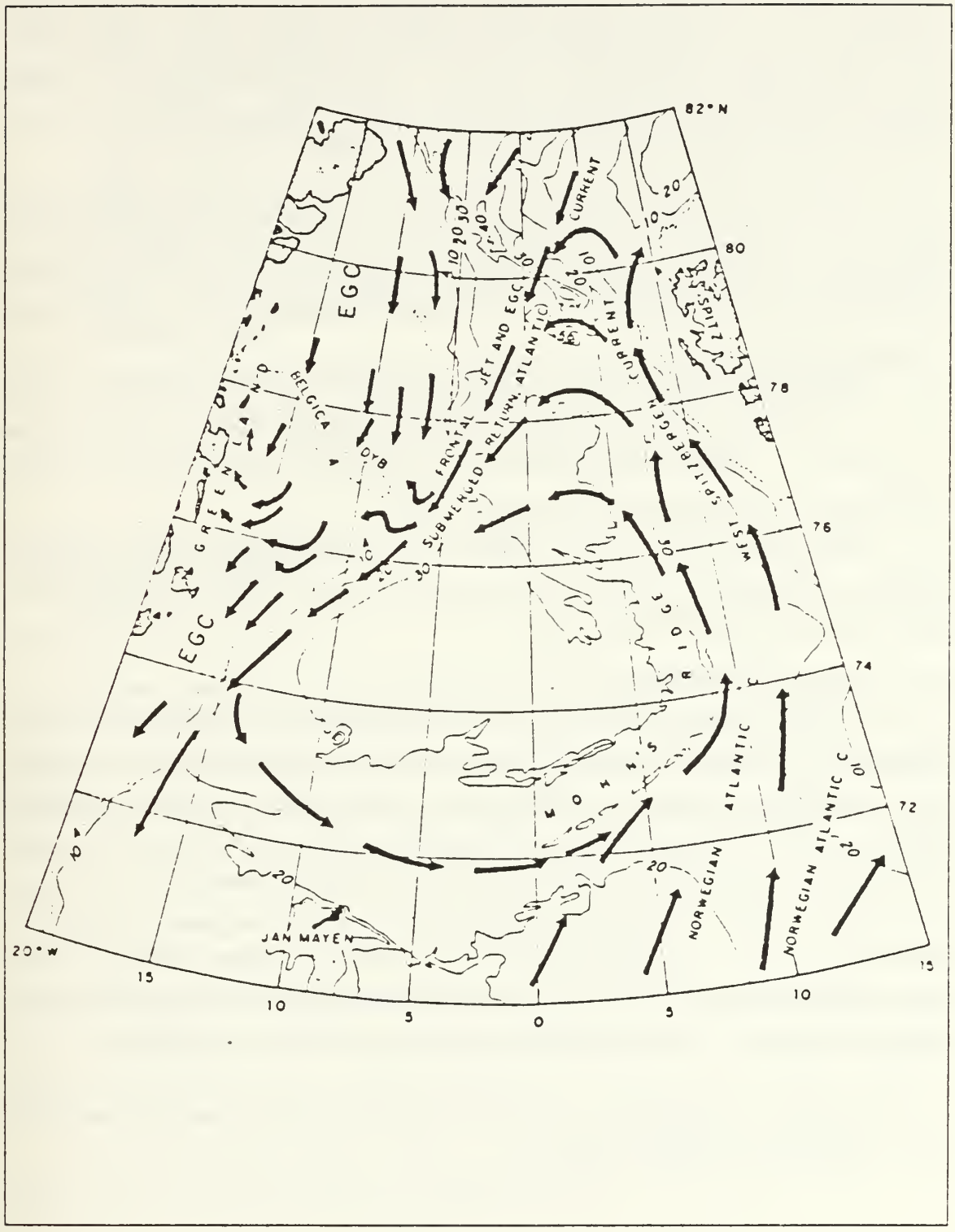


Figure 1.1 A map showing the general bathymetry and circulation in the Greenland Sea (Paquette et al., 1985, p. 4867).

divides into two branches with one turning westward and submerging and then turning southward. This relatively warm water becomes the Return Atlantic Current (RAC). The East Greenland Current (EGC) brings Arctic surface water into the Atlantic Ocean. The cold, fresh EGC contrasts sharply with the warmer and more saline RAC and the boundary between the two gives rise to the EGPF.

Polar Water (PW) extends from the surface to between 150 and 200 m and its temperatures are below 0°C. The surface salinities[1] are often below 30 but increase to about 34.5 at the bottom of the layer. PW originates in the Arctic Ocean but on the Greenland shelf is much modified by processes such as ice melt, freezing, insolation and mixing (Paquette et al, 1985).

Atlantic Intermediate Water is warmer than 0°C and has salinities from 34.5 to 34.9 at about 400 m, remaining fairly constant at greater depths. AIW has upper temperature and salinity limits of 3°C and 34.9, respectively. AIW is found both under the PW and at the surface to the east of the EGPF.

Underlying the AIW at depths below 800 m is the Greenland Sea Deep Water (GSDW). This water is colder than -1°C and has a narrow range of salinity between 34.88 and 34.90 (Aagaard et al, 1985).

## C. SPATIAL COHERENCE

One initial question that is asked about any data set is to what extent it is an organised (or coherent) structure. If such data sets are significantly non-random, it may be possible to interpret or compact them by removing the noise-like components (Mooers, 1985). Oceanographic data is commonly organised spatially and as a result, it can be classified according to its greater or lesser spatial coherence.

There are various methods available to characterise the spatial coherence of oceanographic data. One method is that employed by Monsaigneon (1981) to evaluate the spatial coherence of XBTs acquired in the vicinity of the Maltese Front in the western Ionian Sea using cross-correlation functions. Briefly his method is:

1. Compute a mean temperature profile by averaging all temperature profiles over the data set at specific depths.

[1]Salinity will be reported in the practical salinity scale (gm/kg) as dimensionless quantities (UNESCO, 1981).

2. From a given pair of temperature profiles, compute a cross-correlation coefficient ($\rho_{ij}$). From a set of n profiles, n(n-1)/2 cross-correlation coefficients are generated. The geographical location of each profile in a pair yields a $\nabla X_{ij}$ and a $\nabla Y_{ij}$ defining the horizontal east-west and north-south distance between these two profiles.
3. Relate ($\rho_{ij}$,$\nabla X_{ij}$,$\nabla Y_{ij}$) to the time interval by calculating the distance between two profiles and associating this difference in days between the two profile dates. The cross-correlation coefficients are plotted on a time-distance coordinate system.
4. The sets of values ($\rho_{ij}$,$\nabla X_{ij}$, $\nabla Y_{ij}$) are quantified by 10 km intervals and all $\rho_{ij}$ in a 10 km by 10 km square are averaged to find a single coefficient.
5. These cross-correlation coefficients are plotted and contoured.

A similar method was employed by Brady (1984) to evaluate XBT data acquired from the California Current system.

Using this method, these data were reduced to manageable proportions and the essential structures evaluated.

Another commonly used method for determining spatial coherence is that of empirical orthogonal function (EOF) analysis. EOF analysis (also called principal component analysis, PCA) is used to measure variability (variance) and to characterise (project) the variance onto spatial (or temporal) maps (Preisendorfer, 1982). EOF analysis is somewhat similar to Fourier analysis in that the objective is to represent the original data using orthonormal expansions, solving for the expansion coefficients ($a_n$ and $b_n$). For example, let f(t,x) be a function which represents a mean temperature field in time and space. Then f(t,x) can be represented in terms of M orthonormal functions, $F_k(x_m)$. The objective is to solve for the set of orthonormal functions and their amplitudes, $a_k(t)$, i.e., eigenfunctions and eigenvalues. In spectral analysis $F_k(x_m)$ corresponds to the orthogonal sinusoidal functions. One of the virtues of EOFs over other possible orthonormal expansions is their efficiency of representation.

It is the purpose of this study to investigate a relatively new method for characterising oceanographic data in a detailed manner.

## D. CLUSTER ANALYSIS

The purpose of cluster analysis is to find the 'natural groupings', if any, of a set of individual entities. Cluster analysis allocates a set of entities to a set of mutually exclusive, exhaustive groups such that the entities within a group are similar to one another while individuals in different groups are dissimilar (Chatfield and Collins, 1980). It does this by considering the attributes or characteristics of each entity. In physical oceanography the classical characterisation or 'clustering' of water types is by means of a temperature-salinity (T-S) analysis. Indeed, for entities with only two variables, it is relatively easy to identify the 'clusters', once the data have been plotted in a standard T-S format. However, if additional conservative properties of the water masses are available, such as dissolved oxygen, tritium, nitrate ratios, etc., then plotting beyond two or at most three dimensions is not feasible. One of the aims of this study is to determine if cluster analysis is applicable to the delineation of different oceanographic regimes. Clustering may highlight structure within the water masses that would assist in the optimum deployment of XBTs or sonobuoys.

A cluster can be visualised by considering each entity as a point in n-dimensional space, i.e., the attribute values can be regarded as the coordinates of the entity in attribute space. For example, the geographic position (attribute) of each XBT (entity) can be plotted as a function of latitude and longitude. When plotted and examined, a cluster may be visualised as a region of high density, separated from other dense regions by low density areas. Clusters can be compact, they can be chained or elongated (as in Figure 1.2), or they can assume any other of an infinite number of patterns. Clustering procedures tend to be better at detecting spherical or compact clusters than detecting elongated or serpentine-like clusters (Chatfield and Collins, 1980).

Using different clustering methods with the same set of data will often produce different clustering arrangements or structures. This is because the clustering method imposes its own structure on a data set whether there is any structure there or not. There are situations in which it may not be possible to classify the data set in any useful way and yet a particular clustering method may find structure. It is important to note these considerations before applying cluster analysis.

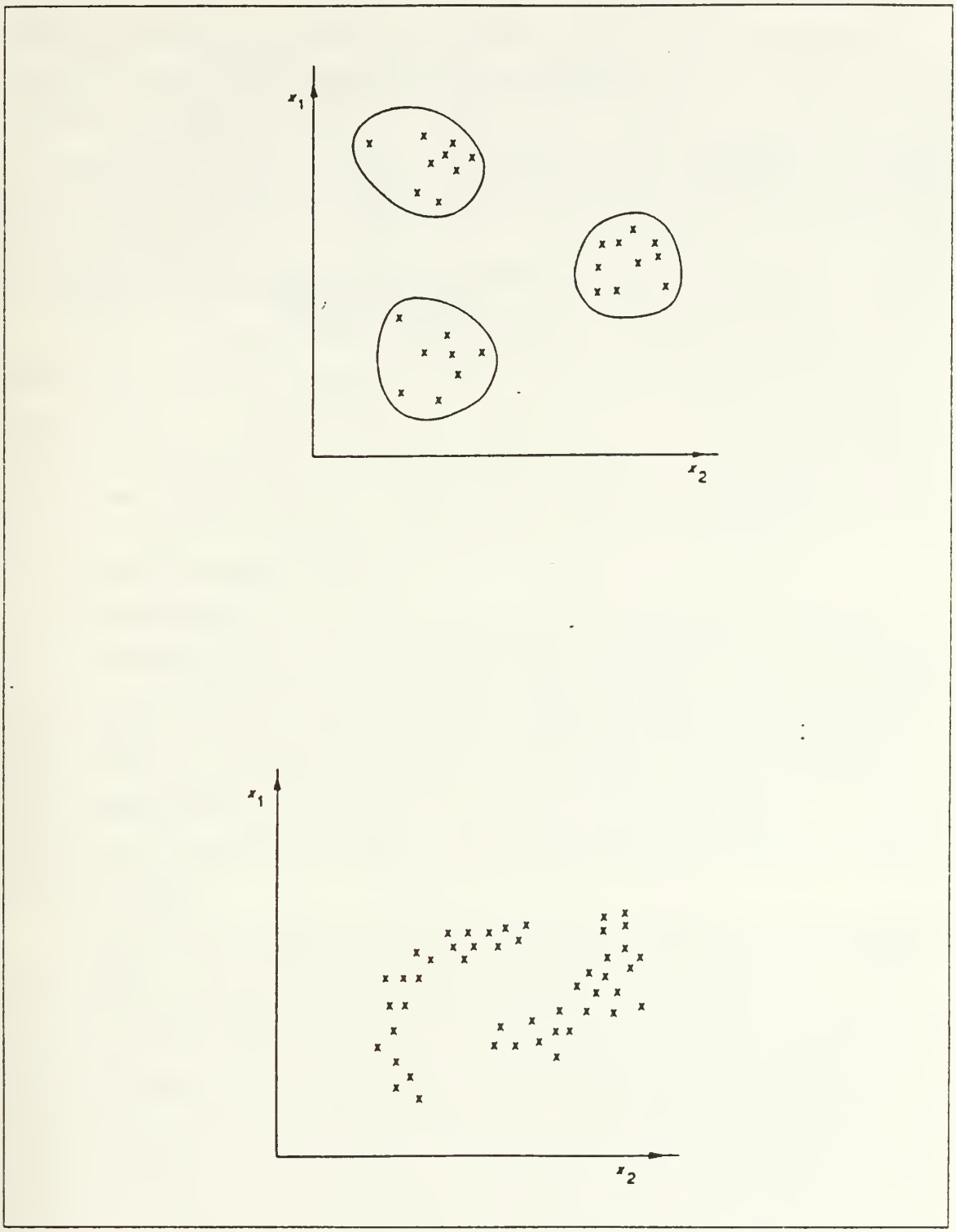


Figure 1.2 A plot of (a) compact clusters, and
(b) elongated clusters (Chatfield and Collins, 1980, p.217).

## E. ACOUSTICAL ANALYSIS

The study concludes with an acoustical analysis of a set of sound speed profiles obtained during the MIZLANT 84 cruise. The aim of this section is to compare and contrast three acoustic models: FACT (Spofford, 1974), RAYMODE (RAYMODE, 1982) and PE (Brock, 1978) models. The first two are range-independent models whereas the latter can accommodate an oceanographic feature such as the EGPF by its ability to process a sequence of sound speed profiles. In addition, the study will investigate whether there is any significant acoustic difference between propagation normal to a frontal feature and propagation oblique to a front.

# II. CLUSTER ANALYSIS

This chapter outlines the details of cluster analysis and explains two particular methods used in this study. It further describes how these methods were applied to two sets of simulated data.

## A. DETAILS OF CLUSTER ANALYSIS

The cluster analysis technique groups entities into subsets on the basis of their similarity across a set of attributes (Lorr, 1983). An entity is an element of the data set and an attribute is a quantitative variable. In this study entities are oceanographic stations and attributes are characteristics of the stations, e. g., temperature and salinity. A cluster then is simply a group of entities whose attributes fall in the same common similarity criterion. For example, a water mass is the cluster one would obtain by classical temperature-salinity analysis.

In applying cluster analysis there are several considerations. An objective method of measuring the similarity (or dissimilarity) of entities is one. Another is to choose the method for forming the clusters. Finally, one must make some initial decision whether or not the entities should be partitioned into separate clusters or be allowed to form a hierarchical or nested arrangement (Lorr, 1983). The first and second considerations are closely connected as described below. The third consideration is often the most difficult. This chapter describes a method to assist in making that decision.

Lorr (1983) lists a sequence of steps that should be considered in a well-designed cluster analysis. A sufficiently large sample of entities must be chosen if the final results are to be meaningful. In this study the data set is 130 oceanographic stations acquired in the East Greenland Current in August and September 1984. The attributes chosen must represent the entities in a meaningful way, for example, temperature and salinity which characterise a water mass. These attributes need to be converted into comparable units for means of comparison. Each set of attributes must be transformed such that the set has a mean of zero and a variance of one. Two distinct clustering algorithms will be examined in this study. In one, the similarity method is chosen independently of the algorithm. In the other, a more sensitive technique, the algorithm is iterative and does not require a predetermined similarity index.

## B. DISTANCE

To obtain a measure of similarity or dissimilarity between clusters, a method of determining the distance between entities (oceanographic stations) is required. In this study the Euclidean metric, or distance measure, is used (Lorr, 1983). The distance between two entities is given by D, where

$$D_{ih} = \sqrt{\sum (X_{ij} - X_{hj})^2}$$

$X_{ij}$ is the value of the attribute j for each entity i, j is a variable of which there are k in number, and i is any entity a,b,...,k,...N. Other metrics can be used, e.g., the congruency coefficient C (Lorr, 1983) given by

$$C = \sum(X_{ij}X_{hj}) / \sqrt{(\sum X_{ij}^2 \sum X_{hj}^2)}$$

This coefficient was considered but the results yielded the same information as those obtained from the distance measure and hence, are not presented here.

Having calculated the distance between each entity or station, the next step is to determine whether it "belongs" or does "not belong" to a cluster. The two cluster methods used in this study differ at this stage and they are considered separately. However, before discussing these two clustering methods the data set to be clustered is described.

## C. THE DATA SET

To further understand the cluster technique, two different sets of artificially constructed (i.e., simulated) data were considered. The first was a set of temperature and salinity pairs each at different locations (Table I). The data shown in Table I were chosen because of the apparent two different temperature regimes from which the data were drawn, warm and cold, and the rather less obvious distinction in salinity regimes. This is similar to that experienced in the East Greenland Current, albeit on a simplified scale.

Although these data are artificial, intuitively one might divide the data into two or three clusters, e.g., {1,2,3} and {4,5,6} or {1,2,3} and {4} and {5,6}.

TABLE I

SIX TEMPERATURE - SALINITY PAIRS
SIMULATED DATA

| | Temperature (°C) | Salinity |
|---|---|---|
| 1. | -1.18 | 31.00 |
| 2. | -1.24 | 31.00 |
| 3. | -1.50 | 32.00 |
| 4. | 2.00 | 32.50 |
| 5. | 2.60 | 33.00 |
| 6. | 2.40 | 33.20 |

The second set consisted of three sets of data produced by a random number generator. Each set contained 50 numbers. The data were normalised with a mean of zero and a variance of one. These data were constructed to provide a test of the clustering algorithms ability to accommodate data sets that had no obvious physical structure, unlike the first set shown in Table I.

The preceding section has outlined the procedure employed in cluster analysis and described the data sets to be used in the analysis. In the following sections each of the two clustering techniques is described and applied to the data sets.

## D. HEURISTIC TECHNIQUE

The first clustering technique is based on an algorithm called, LEADER (Spath, 1980). This algorithm considers each object just once and immediately allocates it to a cluster. A threshold value is first defined to determine if the entity "belongs" to a cluster. The algorithm assigns an entity to a cluster if its distance from the first entity is less than, or equal to, the threshold value.

Heuristic techniques require a suitable threshold value or separation distance to establish which cluster a particular entity should be assigned to. The threshold value can be predetermined or generated within the algorithm. The LEADER algorithm uses a combination of these techniques. Table II shows the input to the LEADER algorithm, using the data of Table I. The number of entities and the number of attributes for each entity are selected. The initial value of the threshold and consequently the value of the incremental step is defined. The original and normalised data are listed in the table. It is the normalised data that are applied to the LEADER algorithm.

TABLE II

INPUT TO THE LEADER ALGORITHM

| | |
|---|---|
| Number of entities | 6 |
| Number of attributes | 2 |
| Threshold value | 0.2 |

Original Data Set

| | | | | | | |
|---|---|---|---|---|---|---|
| Temp(°C) | -1.18 | -1.24 | -1.50 | 2.00 | 2.60 | 2.40 |
| Salinity | 31.00 | 31.00 | 32.00 | 32.50 | 33.00 | 33.20 |

Normalised Data Set

| | | | | | | |
|---|---|---|---|---|---|---|
| Temp(°C) | -0.84 | -0.87 | -1.00 | 0.74 | 1.04 | 0.94 |
| Salinity | -1.16 | -1.16 | -0.12 | 0.39 | 0.92 | 1.12 |

Table III shows how the algorithm subdivides the data into two, three and so on up to six clusters. The algorithm searches for two clusters and then proceeds to search for three and so on up to the number of entities. Each column shows the number of clusters, the separation distance or threshold value and the cluster number to which an entity is assigned. For example, the first line shows a two cluster search with a separation value of 0.2. The first and second entities (T-S pairs) belong to the same cluster, the third to another, and the remaining three are unallocated, as indicated by the zero. The last line shows that when either five or six clusters are allocated the first two pairs belong to the same cluster but all remaining pairs are allocated to unique clusters.

If the algorithm fails to assign every entity to a cluster, it increases the separation distance, RHO, and proceeds again. Theoretically, the algorithm searches for clusters with

$$RHO = J*DELTA \ (J = 1,2,...,JMAX)$$

JMAX being the first J for which all objects are assigned to clusters. Hence, in Table III to assign all the objects to two clusters, a threshold value of 1.2 is required. A key choice is that of a suitable threshold value; it is instructive to examine a plot of threshold value versus numbers of clusters (Figure 2.1). One sees an almost inverse linear relationship between the separation distance and the number of clusters. If two clusters are selected, a fairly large value of RHO is required. In contrast, five clusters can be obtained with a much smaller separation distance of 0.2. In other words, the finer the resolution or distinction between clusters, the shorter the separation distance must be. Clearly then one needs to look carefully at the expected number of clusters and choose an appropriate threshold value. Alternatively, the algorithm can vary the threshold value at will and the results can then be inspected to obtain a "reasonable" number of clusters.

The LEADER algorithm partitions the data set in the same two or three cluster structure that was described above. However, as this data set was small, a further test of the algorithm on a larger data set is described below.

## E. A SECOND SIMULATION

The LEADER clustering technique is now applied to a larger data set. Although the previous example showed that the LEADER algorithm provided results that agreed with an intuitive clustering, it was a small data set. To simulate a more realistic oceanographic situation where there would be upwards of say, fifty stations, three sets of fifty random numbers with no predetermined structure were constructed. Also, it was considered instructive to examine the sensitivity of variations in threshold value to the number of clusters selected, in particular the rate of change in clusters due to incremental changes in threshold value.

Figure 2.2 shows the results of applying the LEADER algorithm to the random data set. One observes that the data can be partitioned into two clusters with a threshold value of 2.6, whereas for a nine-cluster partition the threshold value reduces to 1.0. A 41-cluster partition requires only a threshold value of 0.2.

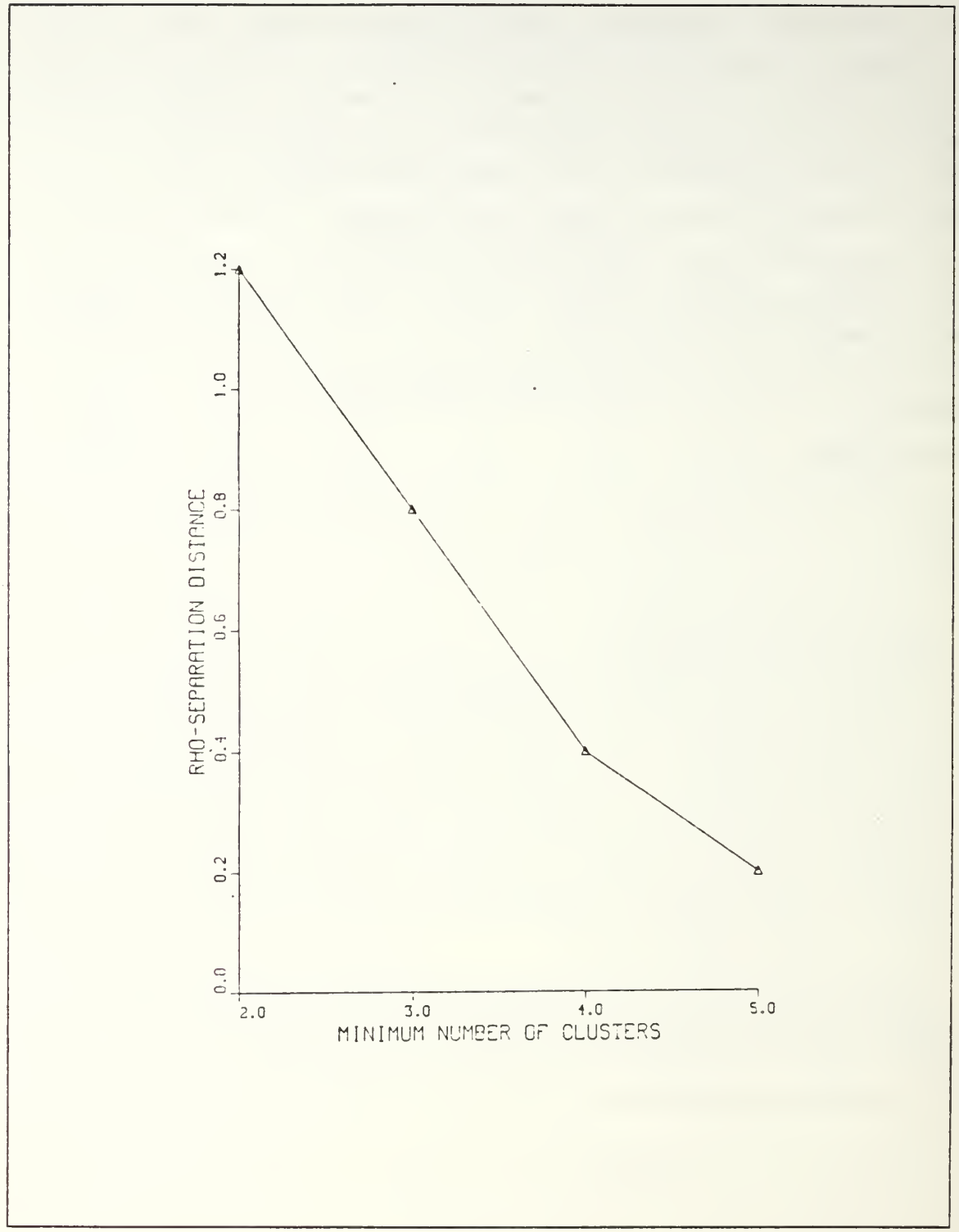


Figure 2.1 A plot of separation distance versus the minimum number of clusters for the simulated data.

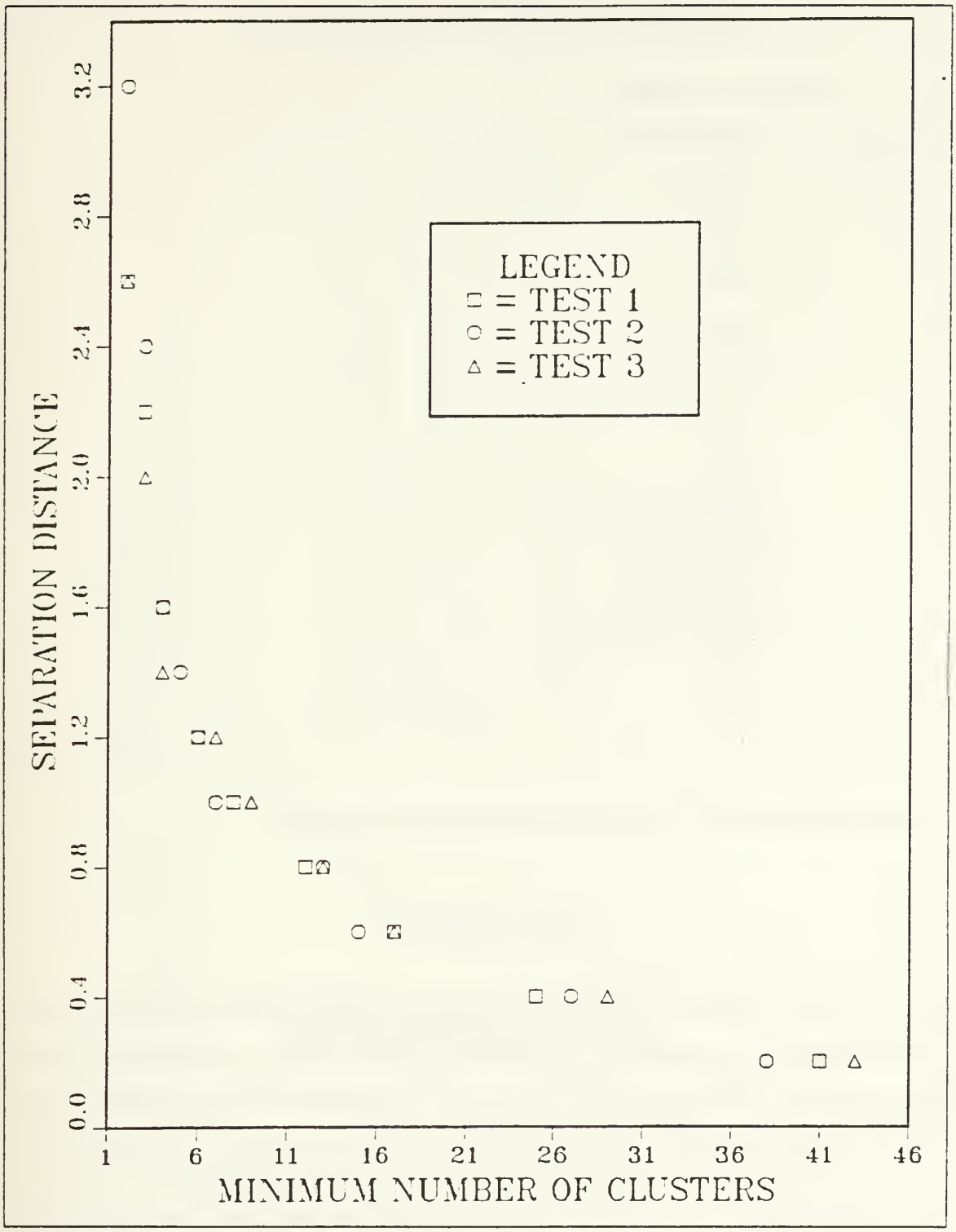


Figure 2.2 A plot of separation distance versus the minimum number of clusters for the three sets of random data.

TABLE III

OUTPUT OF THE LEADER ALGORITHM

| Entity number | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| No. of clusters | Threshold Value | Cluster no. | | | | | |
| 2 | 0.2 | 1 | 1 | 2 | 0 | 0 | 0 |
| 2 | 0.4 | 1 | 1 | 2 | 0 | 0 | 0 |
| 2 | 0.6 | 1 | 1 | 2 | 0 | 0 | 0 |
| 2 | 0.8 | 1 | 1 | 2 | 0 | 0 | 0 |
| 2 | 1.0 | 1 | 1 | 2 | 0 | 0 | 0 |
| 2 | 1.2 | 1 | 1 | 1 | 2 | 2 | 2 |
| 3 | 0.2 | 1 | 1 | 2 | 3 | 0 | 0 |
| 3 | 0.4 | 1 | 1 | 2 | 3 | 0 | 0 |
| 3 | 0.6 | 1 | 1 | 2 | 3 | 0 | 0 |
| 3 | 0.8 | 1 | 1 | 2 | 3 | 3 | 3 |
| 4 | 0.2 | 1 | 1 | 2 | 3 | 4 | 0 |
| 4 | 0.4 | 1 | 1 | 2 | 3 | 4 | 4 |
| 5 | 0.2 | 1 | 1 | 2 | 3 | 4 | 5 |
| 6 | 0.2 | 1 | 1 | 2 | 3 | 4 | 5 |

Of significance here is the shape of the curve, of the form

$$\rho(n) = A\exp(-an)$$

A best fit curve of $\rho(n) = 2.8 \exp(-0.09n)$ (Figure 2.3) fits the data well and, as the data were randomly generated, it is expected that data with no significant natural clustering will also tend to a curve of this type. This suggests the possibility of using this technique for finding the "natural" number of clusters in a data set where there is no initial intuitive "feel" for the number of clusters to expect. Plotting the number of clusters versus separation distance will highlight significant deviations from a curve of the above form. It is likely that this will give some indication of the "natural" number of clusters.

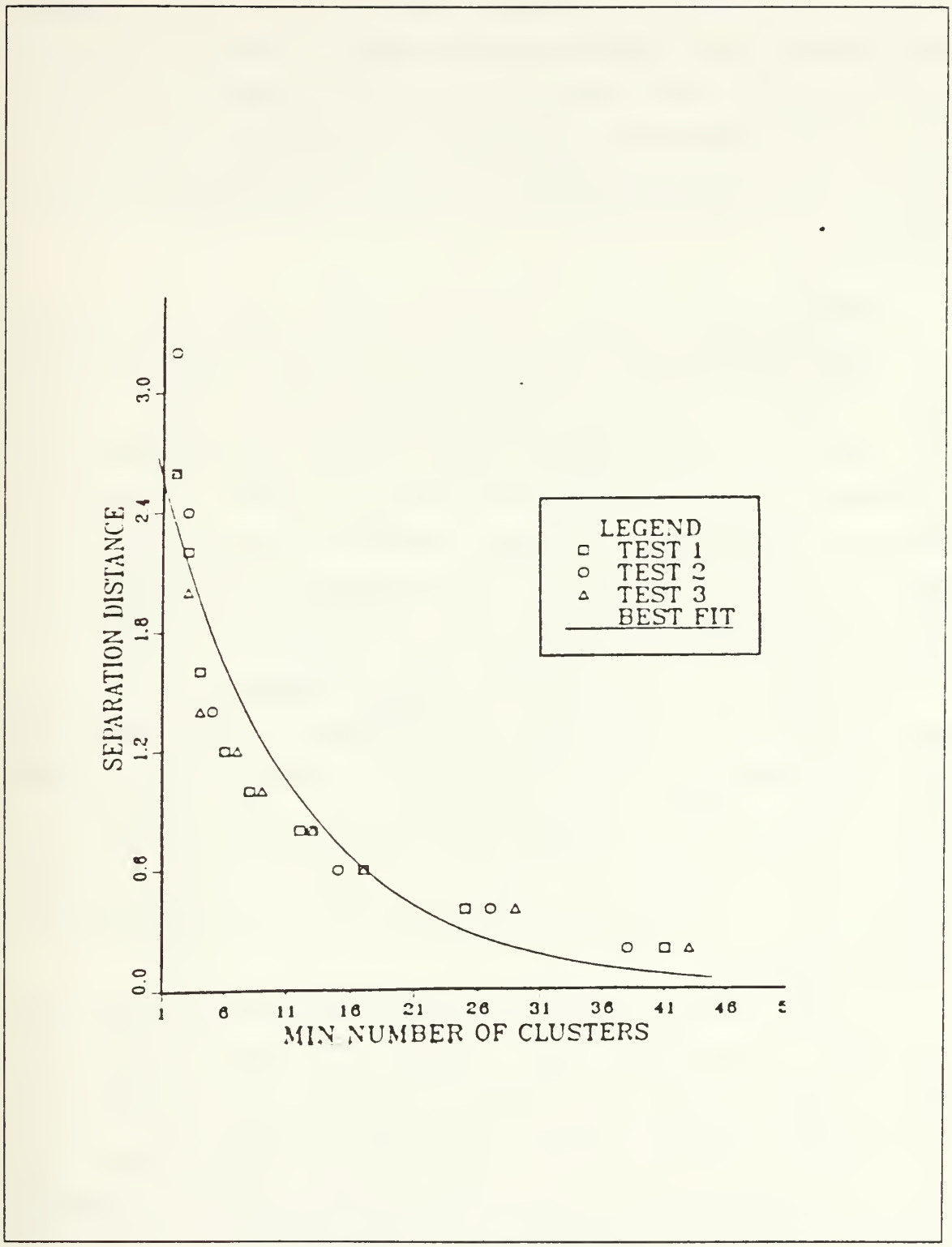


Figure 2.3 A best fit curve for the random data.

An indication of this technique is shown in Figure 2.4. Data that are completely random will tend to plot as a negative exponential curve. If the observed or collected data plot in a different manner, then the 'elbow' of the plot is suggested as the area of interest. Outside of this region, increasing the number of clusters is rather insensitive to the threshold value; or the separation distance is independent of the number of clusters for a regime containing few clusters. Thus, one would examine the clusters in the circled area.

## F. K-MEANS

A more sophisticated clustering technique is the so-called k-means technique (Lorr, 1983). In this technique a sample of size N is sorted or partitioned into k clusters on the basis of the shortest distance between the entity and the k cluster means. The technique works as follows. Initially the data set is arbitrarily partitioned, or a partition from another technique can be used (e.g., LEADER). The centroid of each cluster is then calculated and each entity reassigned to the cluster with the nearest centroid. The sum of the squares of the distances between the members of the J-th cluster and its centroid is denoted by E(J). The k-means technique then minimises D, the sum of the E(J)s, by repeated exchanges of cluster members.

An example of the technique follows (Spath, 1980). Consider the ten points in a plane, as seen in Figure 2.5. These can be intuitively divided into three clusters. As a guide to the technique, consider how the algorithm clusters points 1, 2 and 3. Initially the algorithm is arbitrarily partitioned and each of these points is assigned to a separate cluster. The second iteration combines points 1 and 2 but leaves point 3 in a separate cluster. The third iteration collects the three points into one cluster but also contains points 5 and 8. The remaining iterations first remove point 5 and then point 8 to their final clusters.

The k-means technique is now applied to the simulated data set shown in Table I. The input to the algorithm is similar to Table II and is shown in Table IV. The number of entities and attributes per entity are chosen as are the minimum and maximum number of clusters expected. The user must also decide whether or not to use an arbitrary initial partition; in this case an arbitrary partitioning scheme was selected. As an aid in analysis, each step in the clustering process can be graphically portrayed as in Figure 2.5. The original data and the transformed data are printed as in Table II.

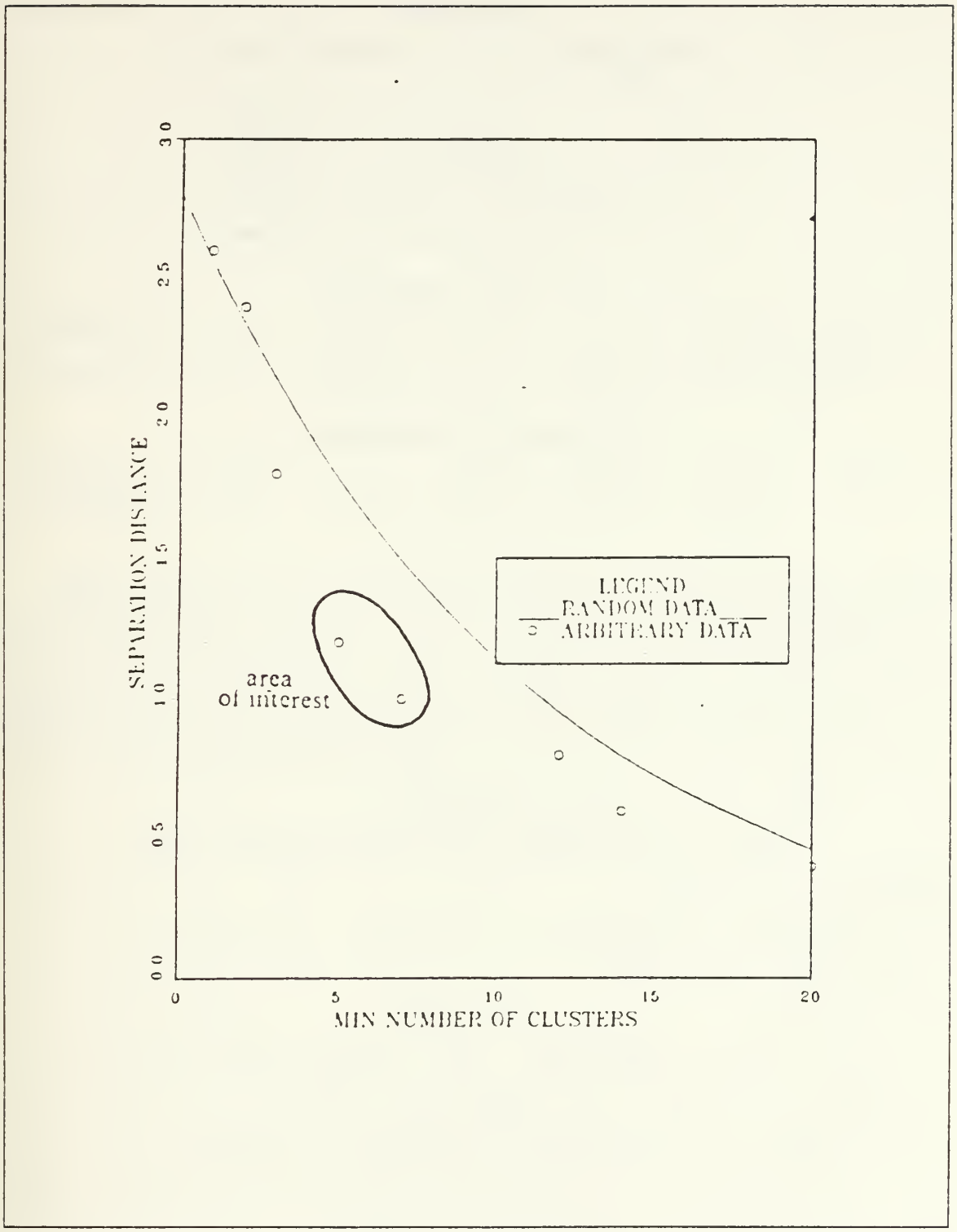


Figure 2.4 An illustration of the area of interest when considering random data.

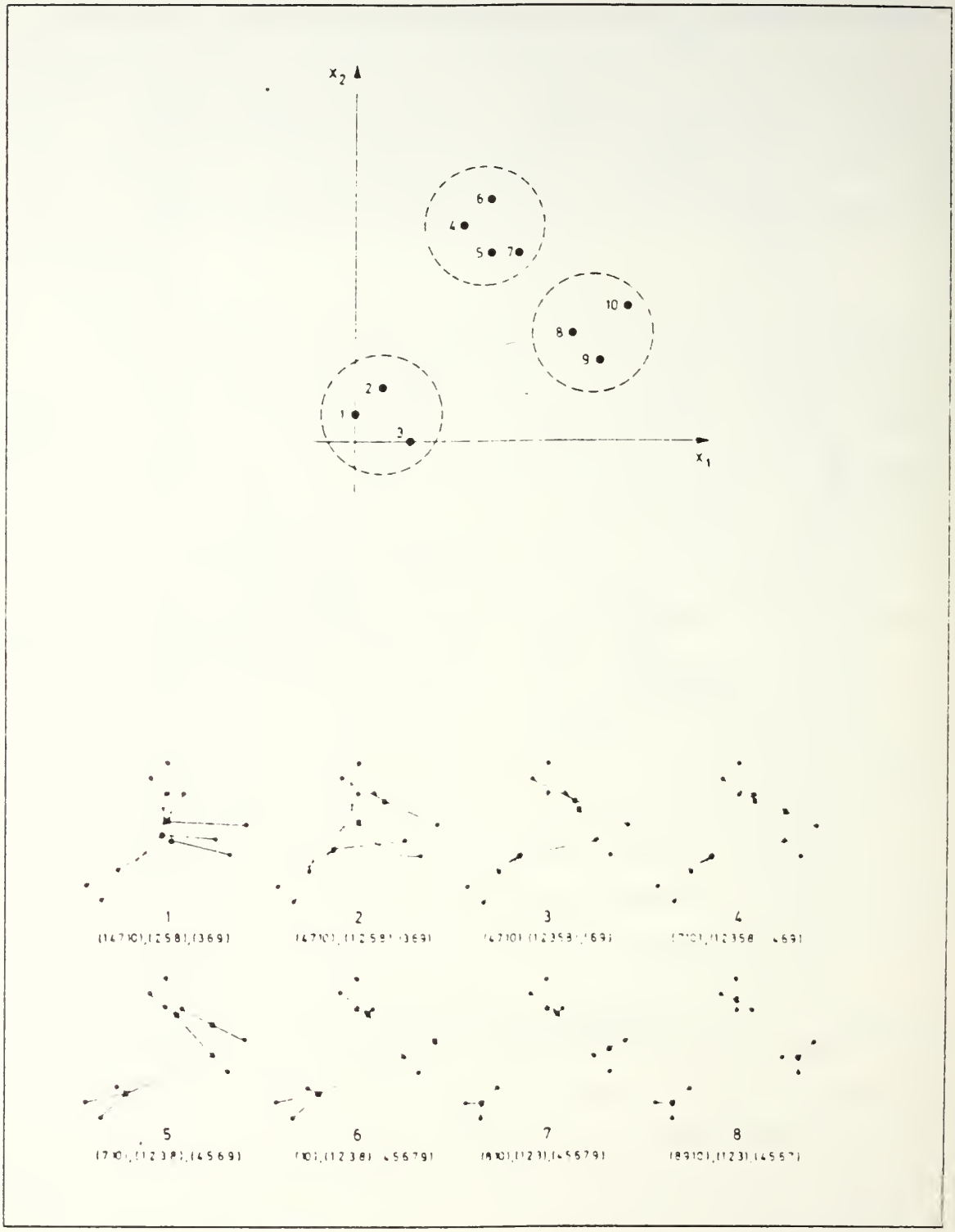


Figure 2.5 Steps in the optimum clustering of 10 data points (Spath, 1980, p. 31).

TABLE IV

INPUT FOR THE K-MEANS TECHNIQUE

| | |
|---|---|
| Number of entities | 6 |
| Number of attributes | 2 |
| Minimum number of clusters | 2 |
| Maximum number of clusters | 6 |

Original Data Set

| | | | | | | |
|---|---|---|---|---|---|---|
| Temp(°C) | -1.18 | -1.24 | -1.50 | 2.00 | 2.60 | 2.40 |
| Salinity | 31.00 | 31.00 | 32.00 | 32.50 | 33.00 | 33.20 |

Normalised Data Set

| | | | | | | |
|---|---|---|---|---|---|---|
| Temp(°C) | -0.84 | -0.87 | -1.00 | 0.74 | 1.04 | 0.94 |
| Salinity | -1.16 | -1.16 | -0.12 | 0.39 | 0.92 | 1.12 |

Table V shows how the technique first arbitrarily partitions the data and then prints out the optimum clustering. In this case the first three stations are allocated to one cluster and the last three to another. In addition, the table shows the centroid locations (-0.9,-0.8) and (0.9,0.8), and the two E(J)s, 0.7 and 0.3, and D, 1.1 (n.b. the E(J)s are shown to one decimal place only).

TABLE V

TWO-CLUSTER OUTPUT FOR K-MEANS TECHNIQUE

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 2 | 1 | 2 | 1 | 2 | Arbitrary clustering |
| 1 | 1 | 1 | 2 | 2 | 2 | Optimum clustering |

Centroids

(-0.9,-0.8) and (0.9,0.8)

Sum of Squares E(J)

0.7 0.3

Centroid Sums D

1.1

The technique, as in the LEADER algorithm, then proceeds to search for three clusters and so on. The results for three clusters are shown in Table VI.

TABLE VI

THREE-CLUSTER OUTPUT FOR K-MEANS TECHNIQUE

```
1     2     3     1     2     3   Arbitrary clustering

2     2     2     1     3     3   Optimum clustering

                  Centroids
   (0.7,0.4)   and  (-0.9,-0.9)  and  (1.0,1.0)

              Sum of squares E(J)
                0.0    0.7    0.0

              Centroid Sums D
                     0.8
```

As can be seen for the two-cluster case, the result is the same as in the LEADER algorithm, but for a three-cluster partition, the k-means technique gives:

1. (-1.18,31.00) (-1.24,31.00) (-1.50,32.00)
2. (2.00,32.5)
3. (2.60,33.0) (2.40,33.20)

whereas the LEADER algorithm gave:

1. (-1.18,31.00) (-1.24,31.00)
2. (-1.50,32.00)
3. (2.00,32.5) (2.60,33.0) (2.40,33.20)

As this was synthetic data, one could justify either clustering scheme. The preceding example illustrates that cluster techniques can partition the data in a "reasonable" manner, and that different techniques may lead to different results. These differences will become more apparent when considering the data from the East Greenland Current.

The k-means algorithm was not applied to a random data set as it is iterative and does not depend on threshold value. The random data set was used to emphasise the importance of threshold values in the heuristic technique.

## G. SUMMARY

This chapter has briefly outlined the concept of cluster analysis and introduced two particular clustering techniques. The two techniques differ in procedure, one a heuristic technique which considers each object only once and then immediately assigns it to a particular cluster. The second technique is iterative and clusters on the basis of the shortest distance between the object and some cluster mean.

Two points should be noted,

1. Cluster analysis can accommodate any number of attributes, i.e., not only temperature and salinity, but, if available, other conservative properties as well.
2. Cluster analysis will always yield a finite number of clusters whether or not there is any "natural" grouping in the data.

The technique introduced here of fitting the number of clusters versus separation distance to a curve of the form

$$\rho(n) = A\exp(-an)$$

is felt to have useful application in determining the natural clustering of a data set.

# III. APPLICATIONS OF CLUSTER ANALYSIS

## A. INTRODUCTION

Before applying the techniques of the previous chapter, the data used in this study are described and a brief summary given of the water masses of the East Greenland Sea. The data set consists of 135 oceanographic stations obtained from the MIZLANT 84 cruise. The data set is shown in Figure 3.1; the EGPF, as defined by classical temperature-salinity analysis, is also shown. The data were obtained during August and September 1984. Full details of the MIZLANT 84 cruise can be found in Bourke and Paquette (1985).

## B. T-S ANALYSIS OF THE EAST GREENLAND CURRENT

The water masses of the area have been described in Chapter 1, as have the characteristics of the EGPF. A brief summary is given here. Polar Water (PW) extends from the surface to 150-200 m and is colder than 0°C. Salinities are less than 30.0 at the surface and increase to about 34.5 at the bottom of the PW layer. Atlantic Intermediate Water (AIW) underlies the PW and at the EGPF is found to the east of it. It is warmer than 0°C with salinities increasing from 34.5 to 34.9 at about 400 m. Greenland Sea Deep Water (GSDW) is found at depths below 800 m; it is colder than -1°C, with salinities between 34.88-34.9.

Stations which show important features of the water masses are plotted on temperature-salinity diagrams (Figures 3.2 and 3.3). Figure 3.2 shows a T-S plot of a station to the east of the front which is in AIW (Station 201, for position see Figure 3.1). Temperatures which characterise this water mass are above 0°C and salinities are above 32.5. A shelf station to the west of the front (Station 247) is also shown. This station exhibits the characteristics of PW. Temperatures are below 0°C and salinities are between 30.0 and 34.5. A third station (210), located in the frontal mixing zone exhibits characteristics of both PW and AIW underlying the PW at depth.

Figure 3.3 highlights some relatively subtle features of the water masses over the continental shelf. The T-S plot suggests that the shelf waters could be divided into two regimes. The water to the west is cold and fresh whereas the water to the east is equally cold but more saline. One sees in this figure that the waters to the west, as typified by Station 225, have a relatively smooth progression from PW to AIW. The

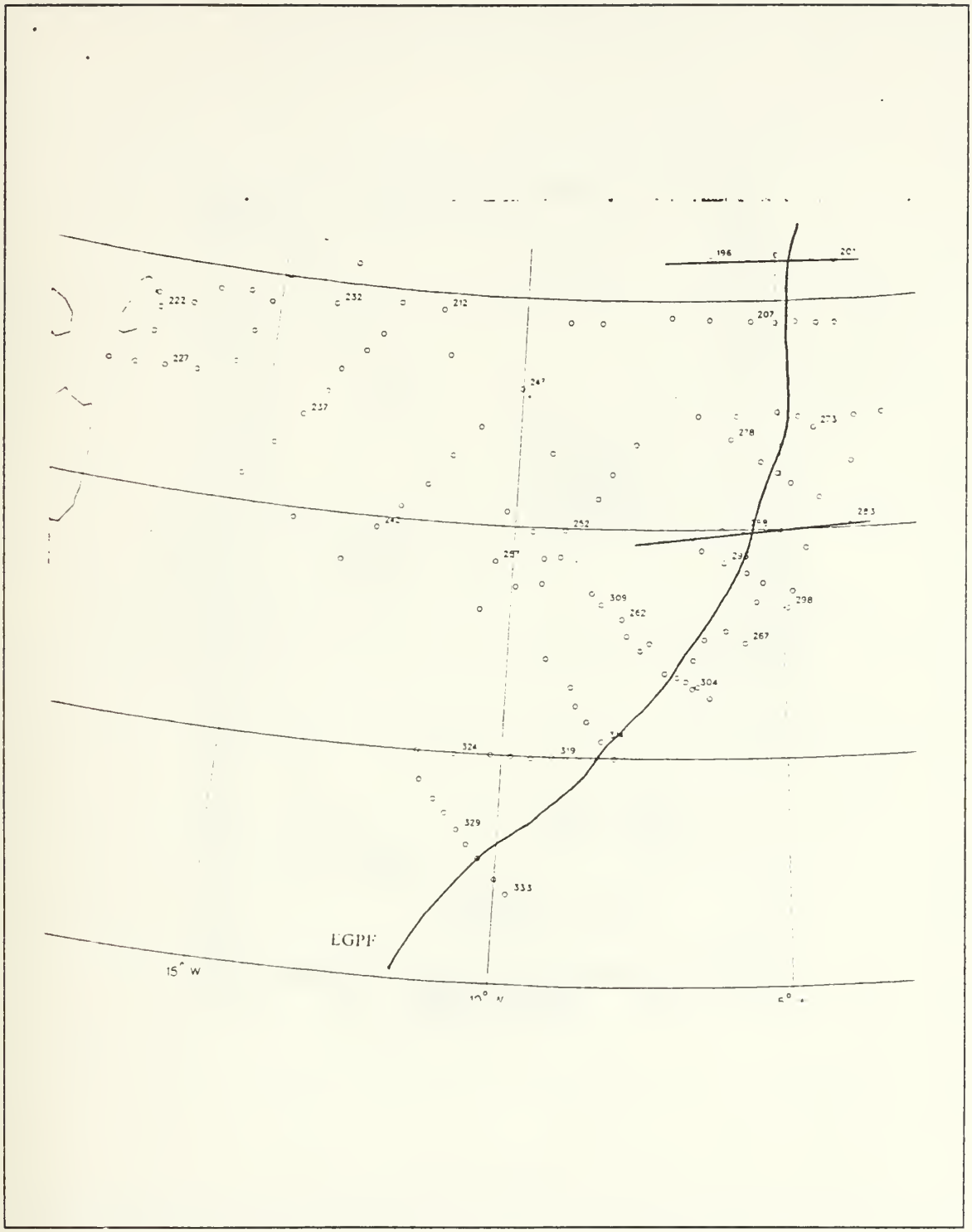


Figure 3.1 A map of the oceanographic stations forming the data set. Two frontal transects used in the cluster analysis are indicated. The EGPF is also shown.

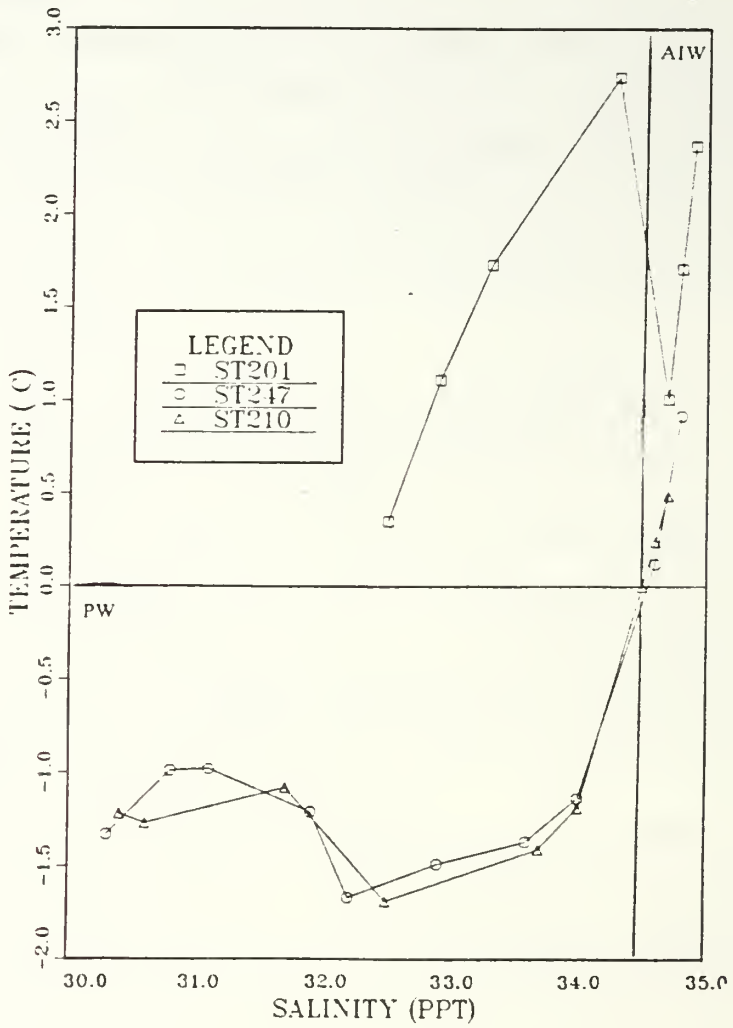


Figure 3.2 A T-S plot of an AIW station (201) to the east of the EGPF, a PW station (247) on the shelf and a further PW station (210) in the frontal region.

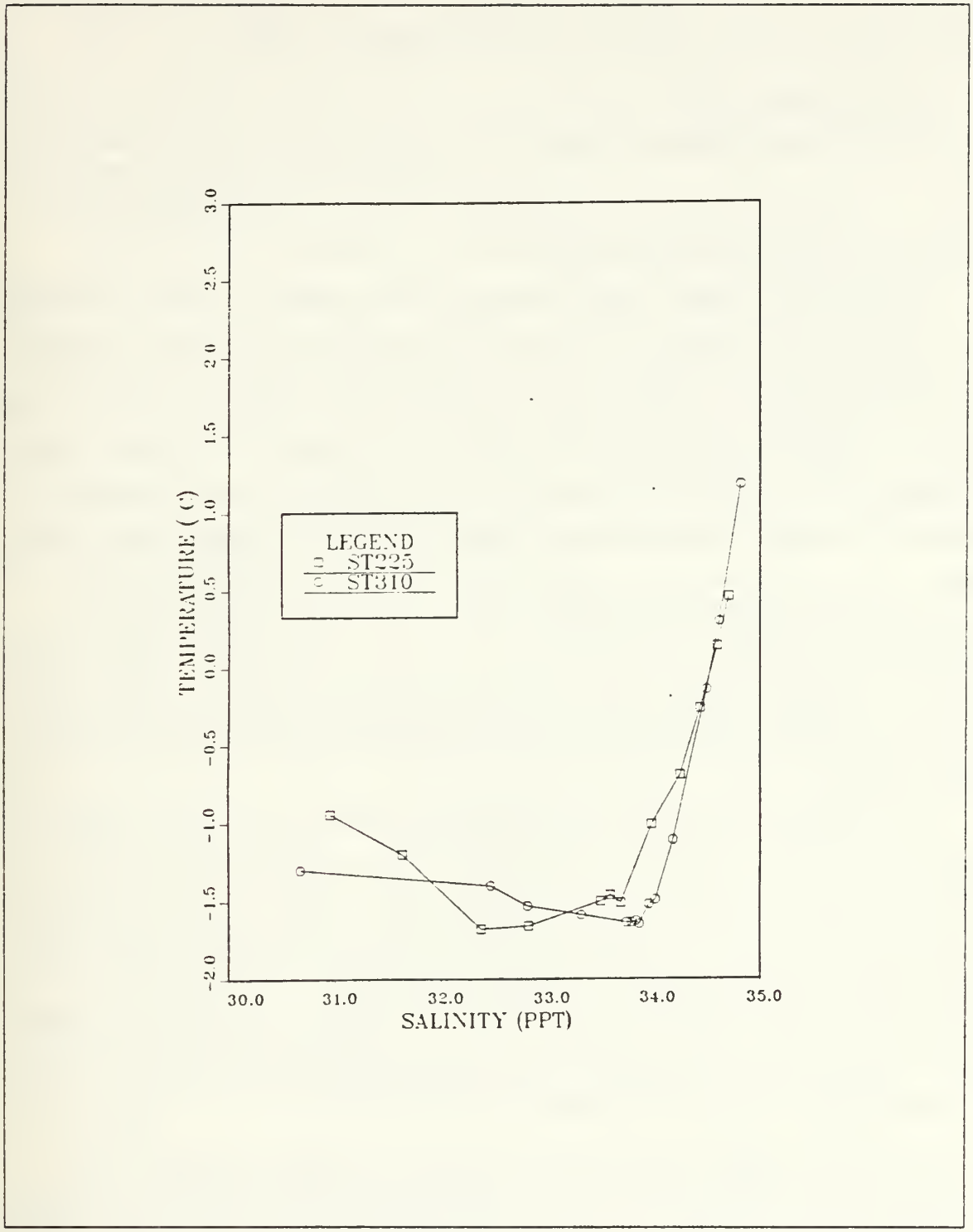


Figure 3.3 A T-S plot of two shelf stations.
Station 225 is situated on the western part of the shelf.
Station 310 is situated to the east of the shelf.

easterly waters have a more discontinous 'jump' to AIW as shown by the T-S plot of Station 310, in the salinity range 34.75 to 35.25 (Paquette et al, 1985).

## C. CLUSTER ANALYSIS

Unlike classical temperature-salinity (T-S) analysis which is used to characterise a water mass on the basis of its temperature and salinity, cluster analysis looks for 'natural' groupings in such data. The technique groups the data into clusters which may be used to characterise water masses.

Temperature and salinity values were examined for different subsets of the data set. Initially all the stations were considered using temperature and salinity values at two different depths. An average value of temperature and salinity for the 10-20 m layer was considered as a representative sample for the surface mixed layer. In addition, temperature and salinity at 150 m was considered. A smaller subset consisting of stations in the vicinity of the front was also examined. This section uses values from the 10-20 m layer and also a single value at 500 m depth. As described earlier cluster analysis can accommodate several variables (attributes). With this in mind, two frontal transects were examined (Figure 3.1). The frontal transects were examined using different combinations of the attributes; temperature, salinity and location. In addition to the above, cluster analysis was used to investigate the 'warmer' stations to the east of the front by following temperature along a constant density surface.

The reason for the selection of observations between 10-20 m was to obtain a measure of the surface mixed-layer that is free from local surface affects such as melting ice which could distort surface temperatures and salinities. A depth of 150 m was chosen because the zero degree isotherm is found at this depth over much of the continental shelf. This isotherm which marks the boundary between PW and AIW. A depth of 500 m, which is in the deep water below the thermocline, is chosen to provide a sensitivity test for the cluster technique, since at this depth, one might expect only one cluster, based on classical T-S analysis, as seen in Figure 3.2.

The EGPF divides the study area into two water masses and cluster techniques are initially used to demonstrate such a characterisation. The front occupies a band of about 40-60 km in width and it is reasonable to assume that there is a transition region with groups of stations which do not clearly belong to the cold, fresh or warm, saline water masses. Thus, in addition to two clusters, the possibility of three (or more) clusters is also considered.

## D. ENTIRE DATA SET

Cluster techniques are first used to analyse the entire data set. The purpose of applying the techniques to all of the data is to examine how cluster analysis deals with a relatively large set of entities that already have a clearly defined distinction between water masses to provide a good comparison.

### 1. The mixed layer

Both clustering techniques divide the data in similar ways with the k-means algorithm resolving the front somewhat better than that of the heuristic algorithm (Figure 3.4). The heuristic technique places the cluster boundary to the east of the EGPF, particularly in the northern part of the area, whereas the cluster boundary determined from the k-means technique is more closely aligned with the EGPF. The heuristic technique also depends on the starting point. As such, there appears to be some 'inertia' before the technique is able to resolve a new cluster. The k-means technique more accurately defines the boundary separating the PW and the AIW most likely because of the iterative nature of the technique.

Both techniques, when used for a three cluster search, identify the EGPF as a natural division (Figures 3.5 and 3.6). There is, however, a significant difference between the two. The heuristic technique splits the data into two main sets and only a handful of stations are found in the third set, with the main division following the front exactly. The k-means technique divides the group into three distinct clusters. The warm water to the east of the front compromises the first cluster; the cold water to the west is divided into two clusters. The cold, fresh water over much of the shelf and the cold but slightly more saline water immediately to the west of the front are distinguished by this iterative technique. The division of the cold water mass into two parcels is interesting in that this result directly parallels that obtained by using a classical T-S analysis for the same data, Figure 3.3. The region is homogenous in temperature but changes in salinity by one part per thousand from west to east. The k-means technique makes this subtle distinction here, showing its greater sensitivity.

### 2. 150 m

As outlined in Chapter 1, the polar front slopes towards the west with increasing depth. Thus, a two-cluster search at 150 m is expected to yield similar results to that at the surface but with the contour lines displaced westward. This is the case, as can be seen in Figure 3.7. A three-cluster search at 150 m also shows similar results to those at the surface (Figures 3.8 and 3.9). The heuristic algorithm groups the

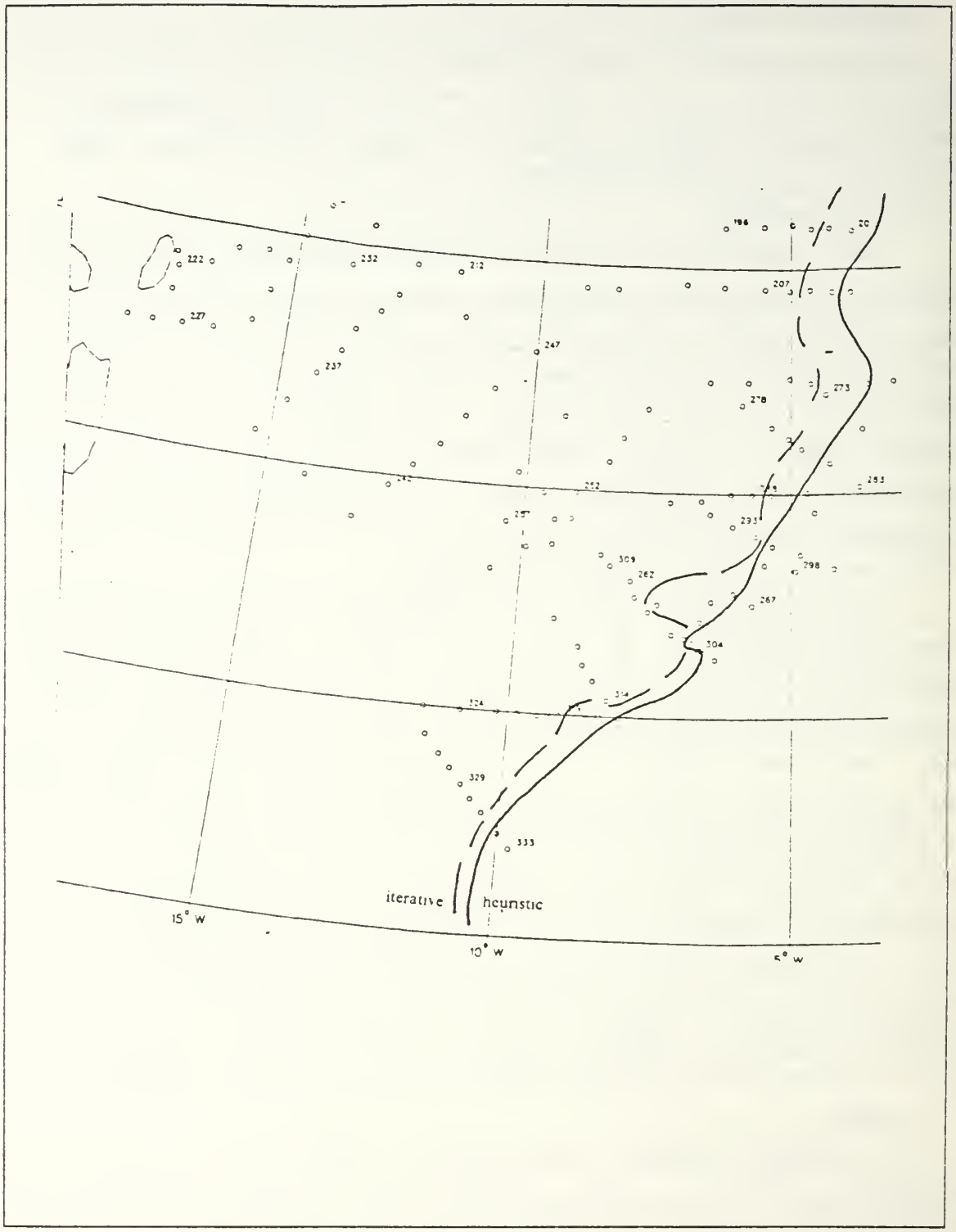


Figure 3.4 Results of a two-cluster search in the mixed layer using the heuristic and iterative techniques.

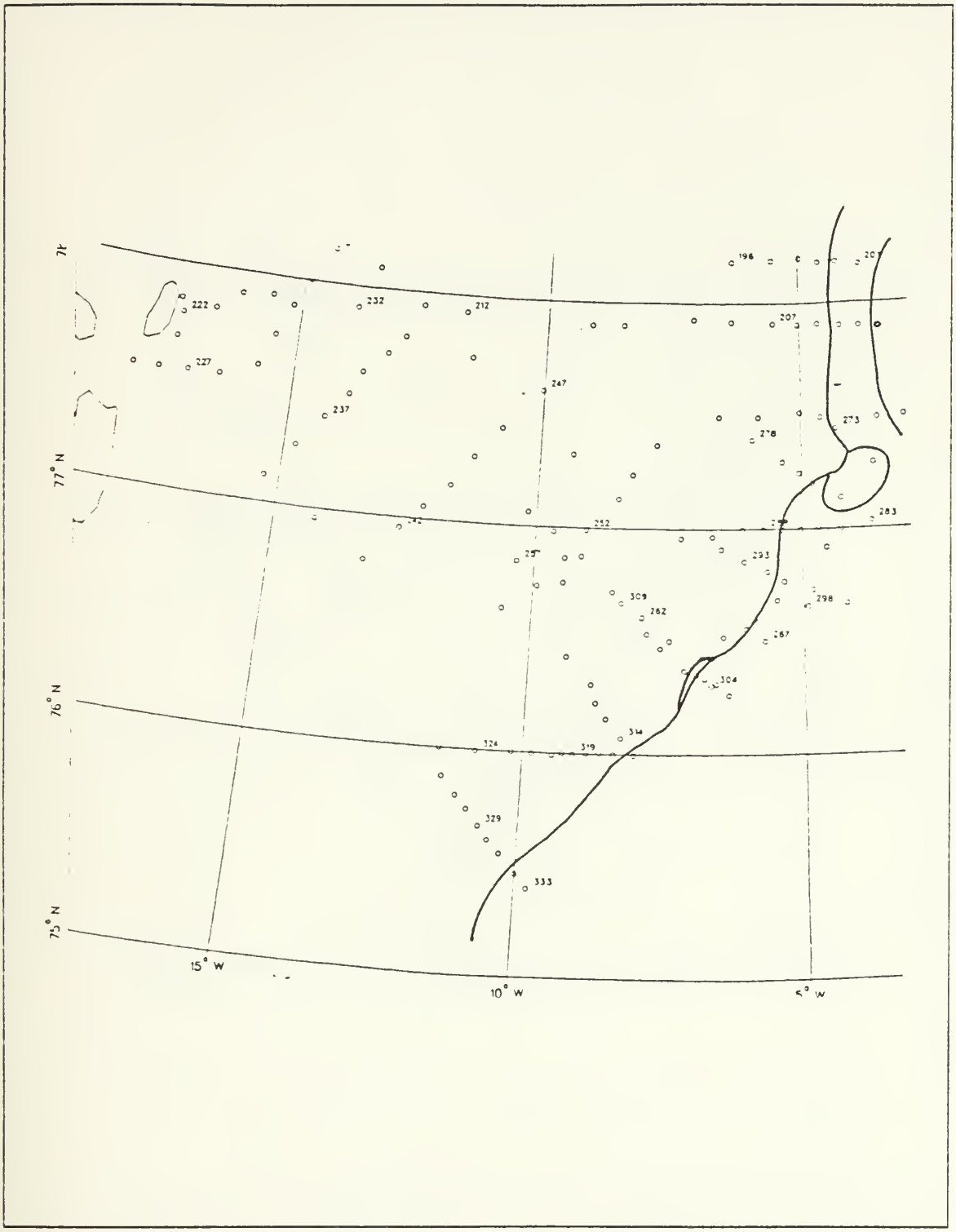


Figure 3.5 Results of a three-cluster search in the mixed layer using the heuristic technique.

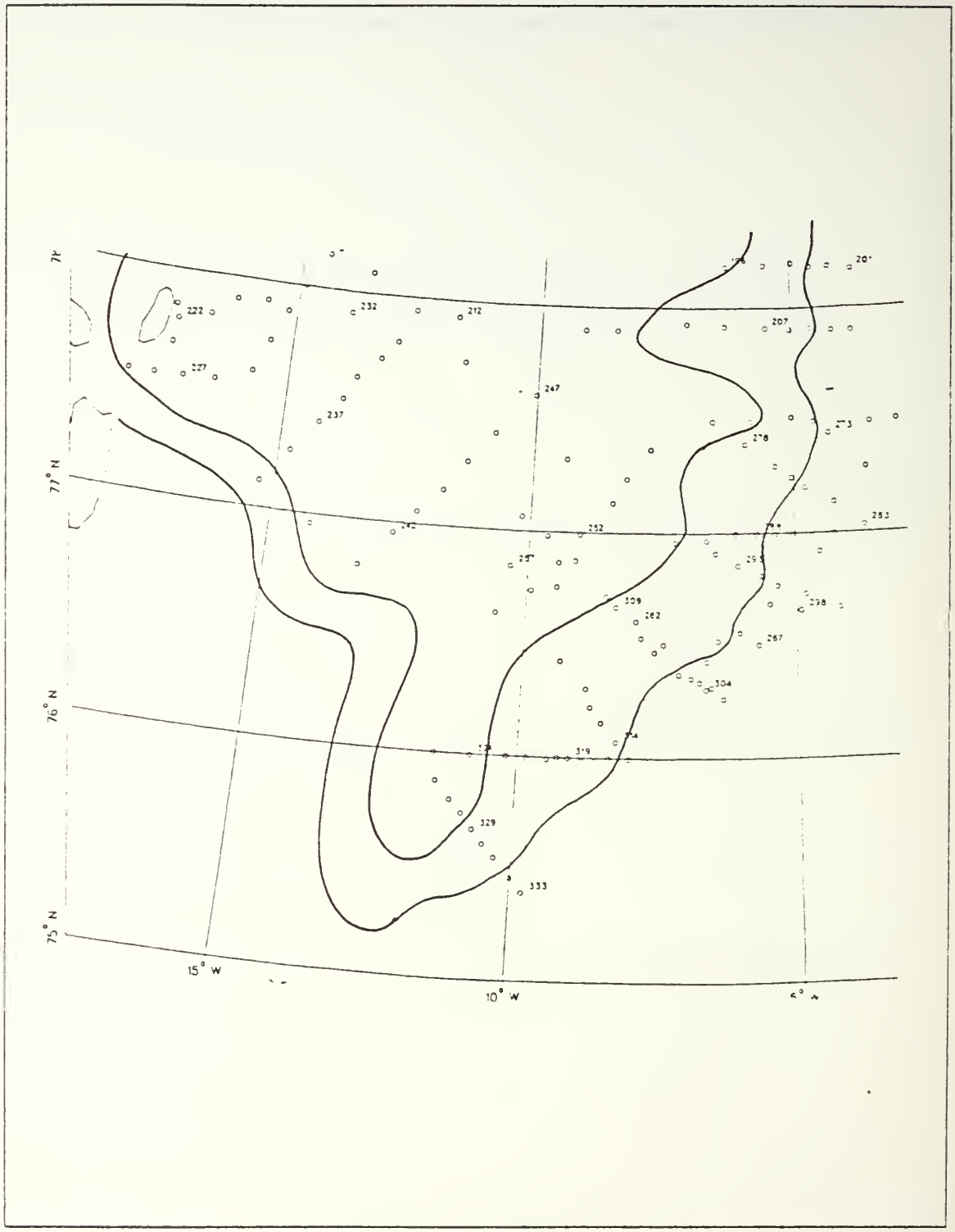


Figure 3.6 Results of a three-cluster search in the mixed layer using the iterative technique.

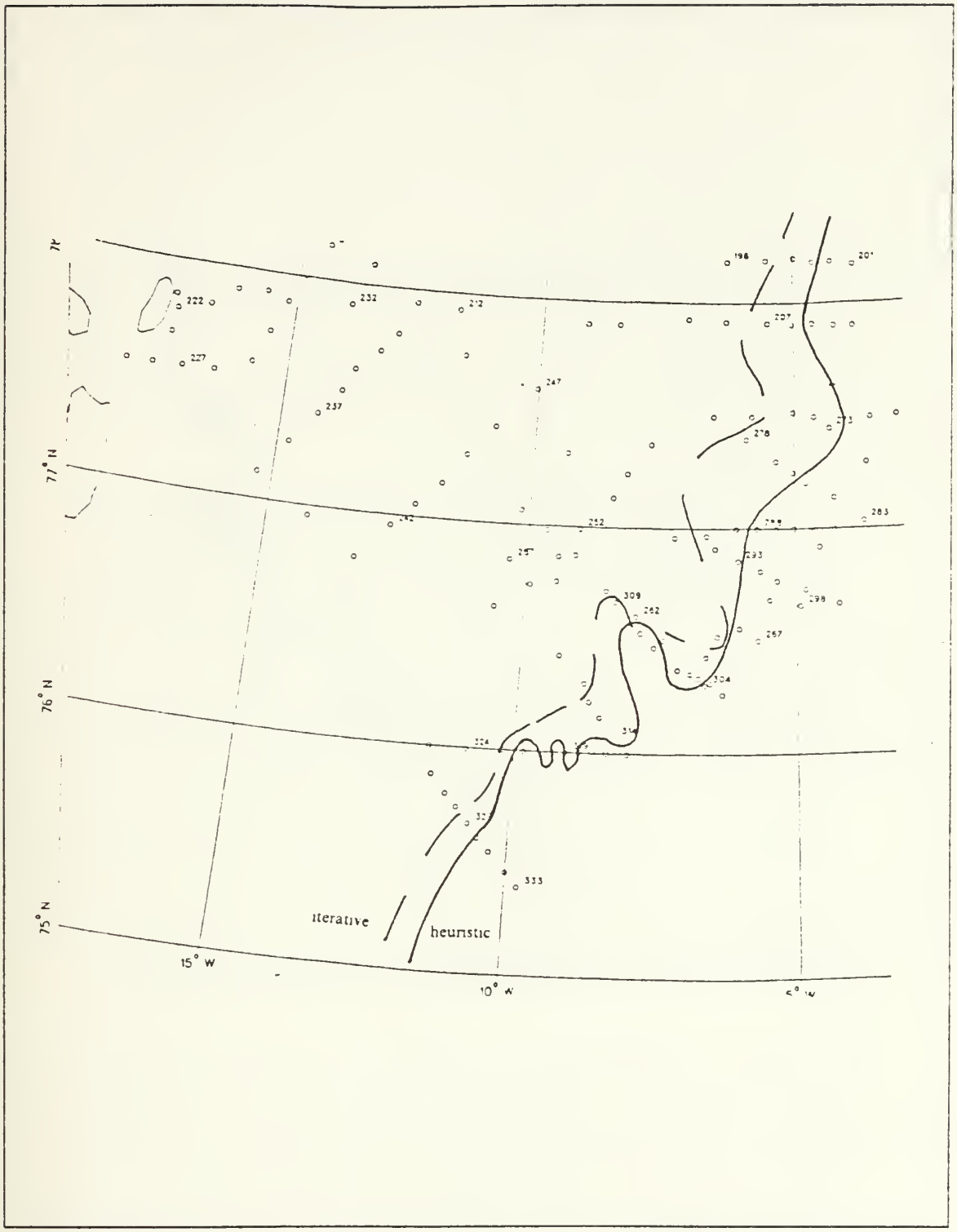


Figure 3.7 Results of two-cluster search at 150 m using the heuristic and iterative techniques.

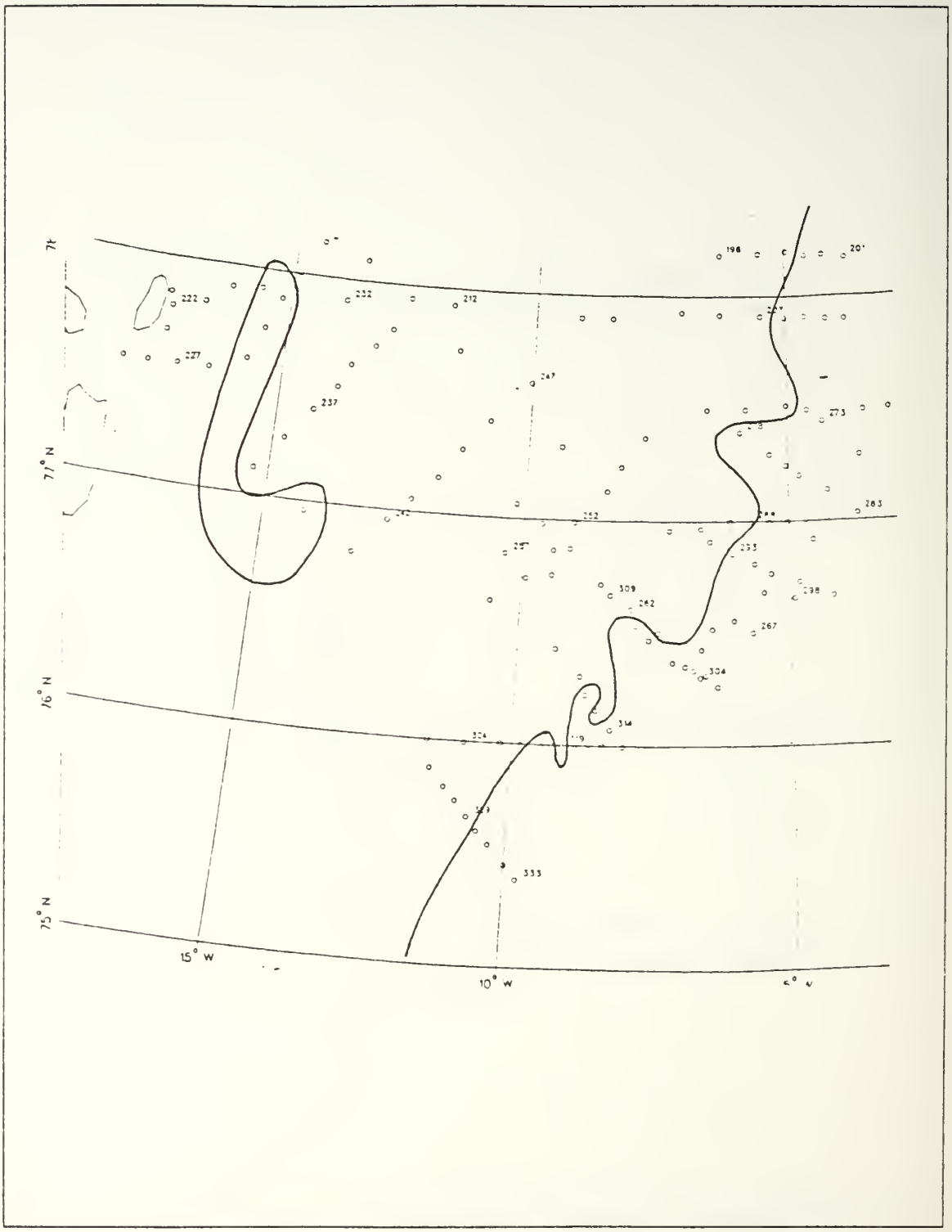


Figure 3.8 Results of a three-cluster search at 150 m using the heuristic technique.

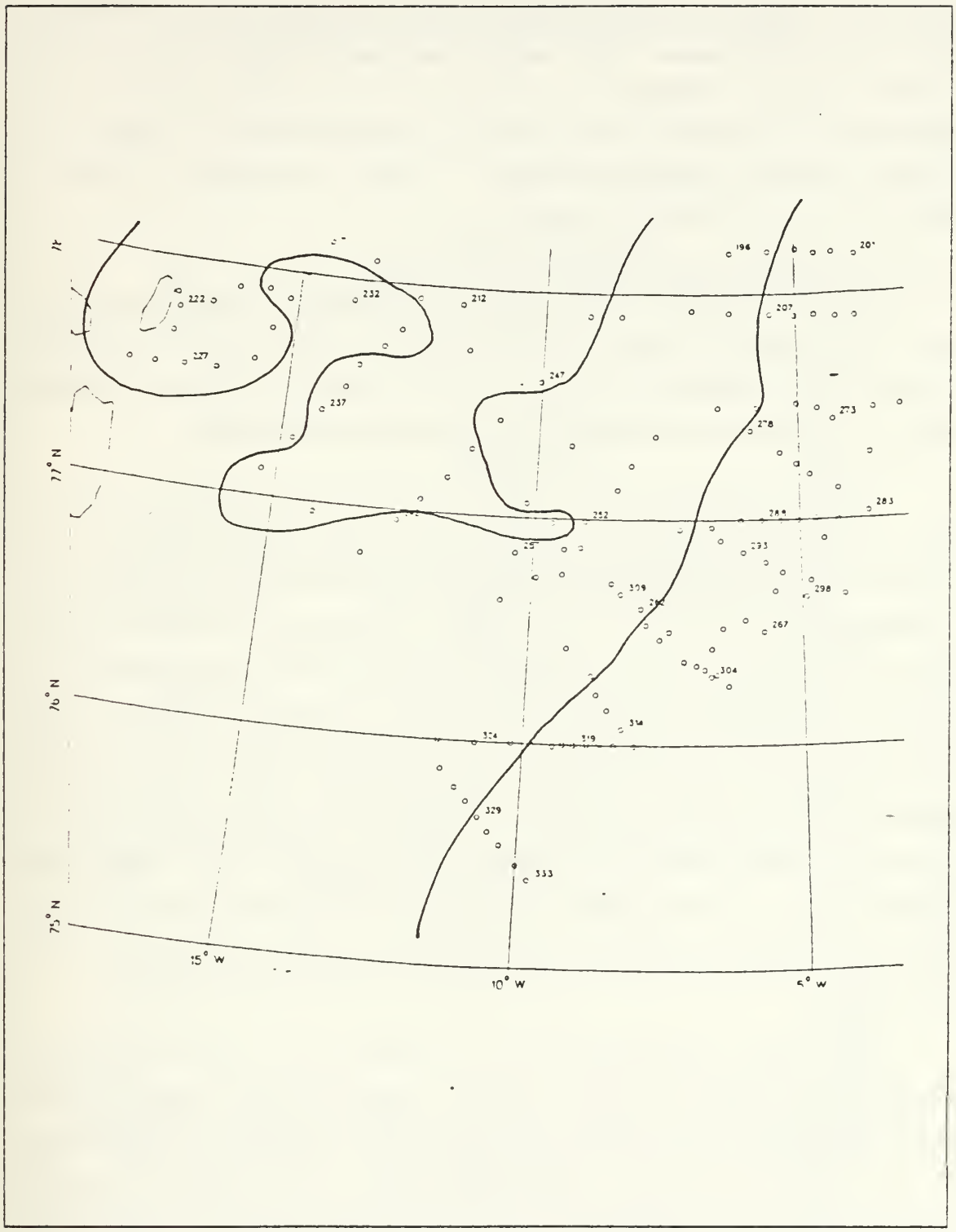


Figure 3.9 Results of a three-cluster search at 150 m using the iterative technique.

data into two clusters with a few outliers, whereas the k-means technique divides the cold water into two clusters. These results again show the greater sensitivity of the k-means technique. Intuitively there are either two clusters, PW and AIW, or as explained above, three clusters with the cold water being divided into two regimes. The heuristic technique cannot accommodate this subtlety and allocates a handful of stations to a cluster with little physical basis.

## E. STATIONS IN THE VICINITY OF THE FRONT

A subset of the data base was constructed for 40 stations in the vicinity of the EGPF. These data were collected over a ten-day period and are more synoptic than the whole data set which was collected over 24 days. Classical oceanographic analysis indicates the front divides this data subset into two groups of equal size.

### 1. 10-20 m

Although these data have previously been examined as part of a larger data set (see above), it was hoped, by examining a smaller, more synoptic data set, to avoid the 'inertia' problem mentioned earlier. However, this is not the case, possibly because the starting points of both sets are the same. Examining the two-cluster searches, Figure 3.10, one sees that neither technique corresponds exactly with the classical T-S analysis. The three-cluster searches, Figures 3.11 and 3.12, show the utility of cluster techniques, in that they depict the front as a horizontal band in the ocean. The heuristic technique, however, provides slightly ambiguous results (Figure 3.11). It could be argued that it provides two water mass clusters with a few warmer outliers or, that there is a broad transition zone mainly to the east of the front. The k-means technique (Figure 3.12) also suggests a broad transition zone again mostly to the east. One also sees that the transition zone straddles the front and that the warm cluster regime is better defined than the few outliers defined by the heuristic technique.

### 2. 500 m

The purpose of analysing the data at 500 m was to test if either cluster technique could simulate the natural structure, given that cluster techniques often identify clusters even in the absence of natural clusters. Intuitively one might expect the 500 m data to reveal one cluster with perhaps a few outliers, as described previously. That clustering may impose artificial structures is apparent in Figure 3.13. The heuristic technique ( Figure 3.13) suggests a result that is similar to the 'intuitive' case in that it clusters the set into one regime with just two outliers. The tortuous

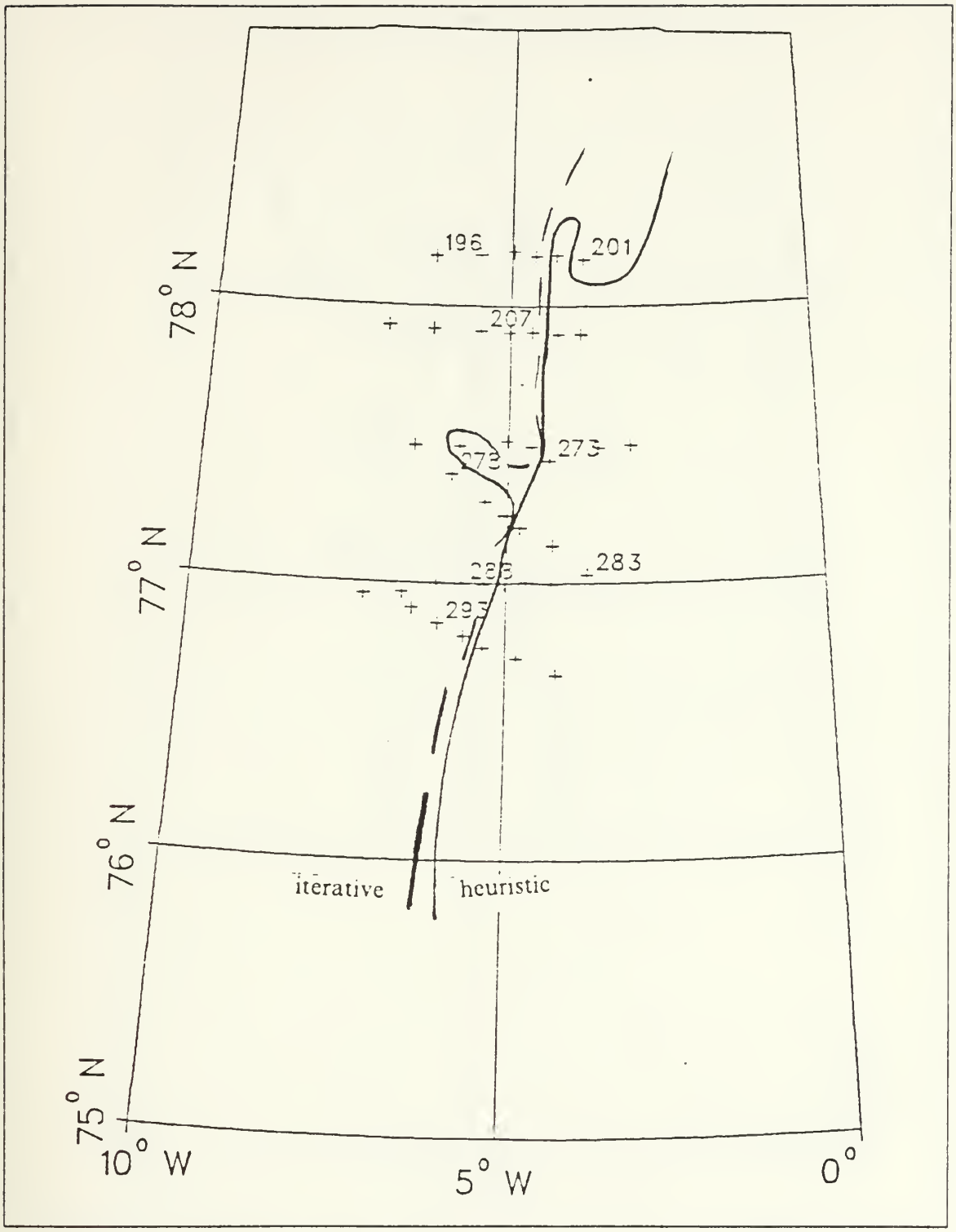


Figure 3.10 Results of a two-cluster search in the mixed layer for Frontal Stations using the heuristic and iterative techniques.

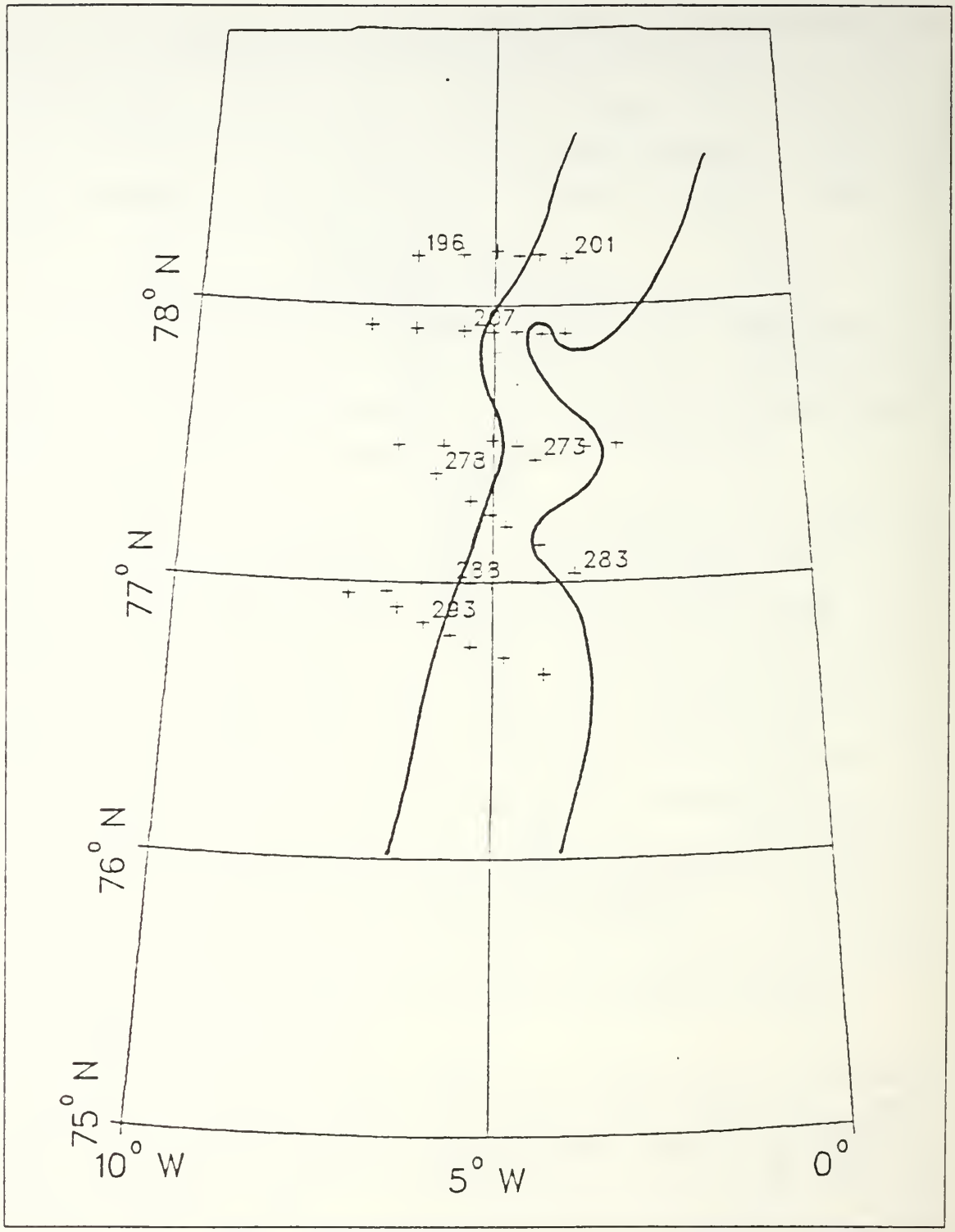


Figure 3.11 Results of a three-cluster search in the mixed layer for Frontal Stations using the heuristic technique.

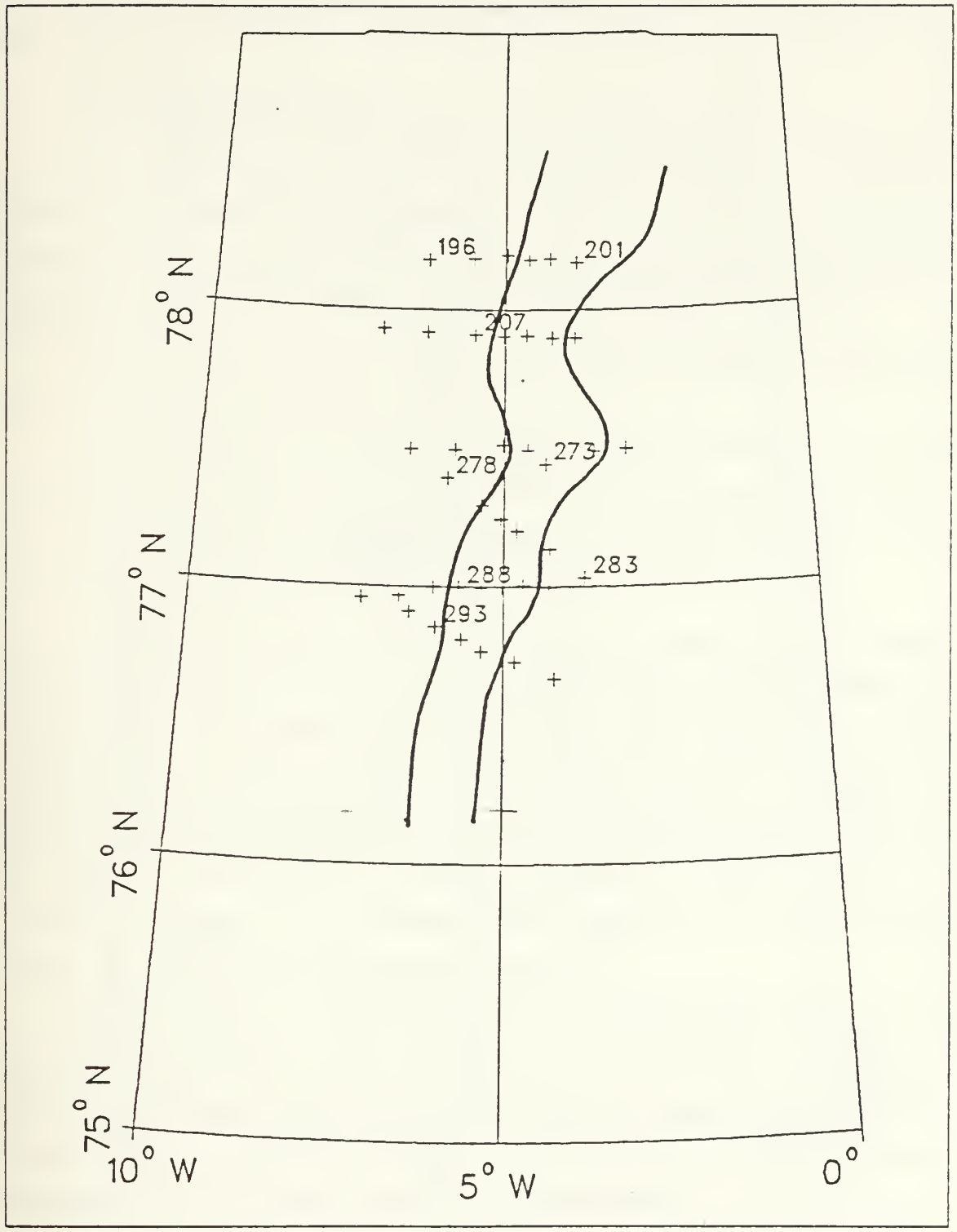


Figure 3.12 Results of a three-cluster search in the mixed layer for Frontal Stations using the iterative technique.

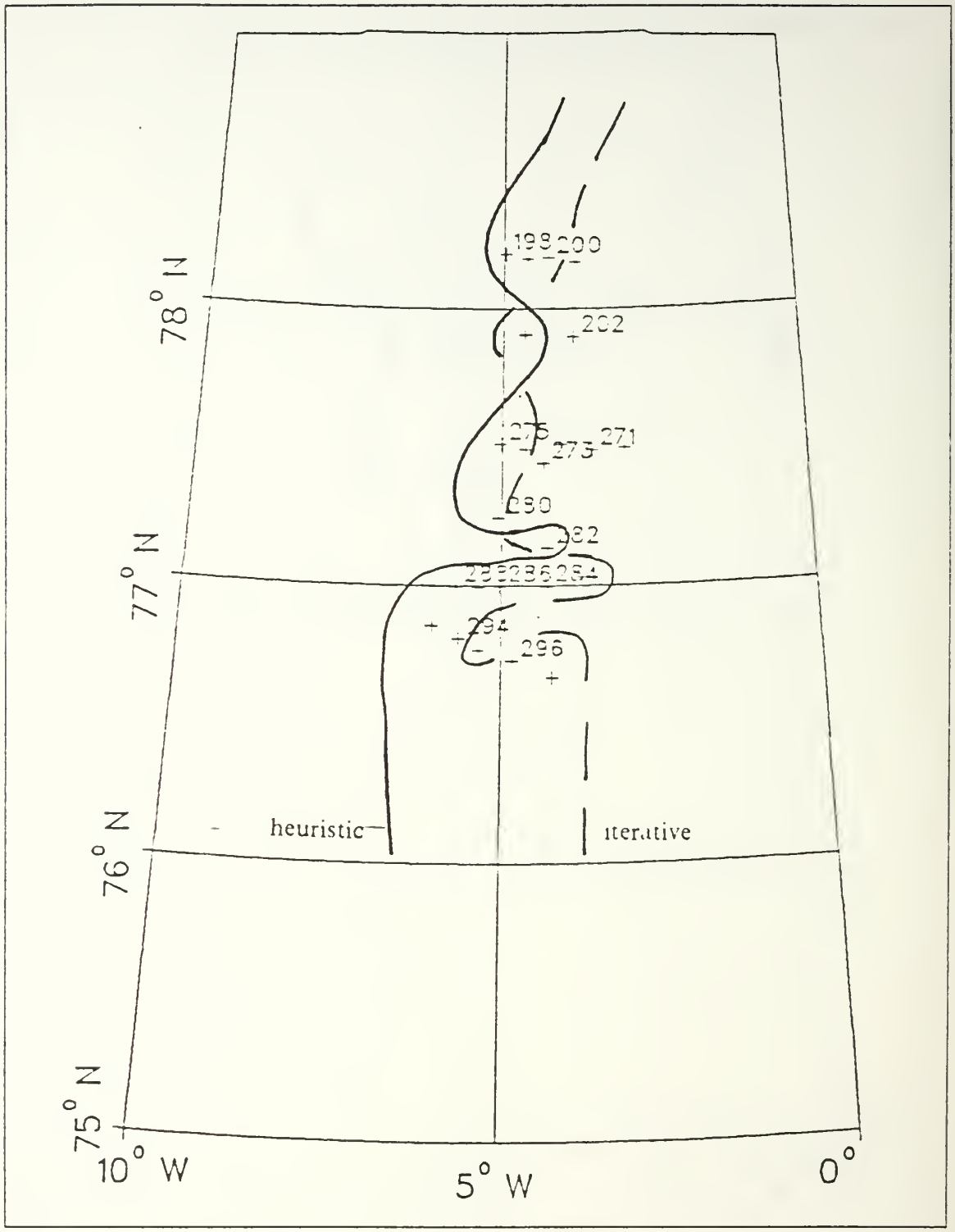


Figure 3.13 Results of a two-cluster search at 500 m for Frontal Stations using the heuristic and iterative techniques.

contours of the three-cluster searches, Figures 3.14 and 3.15, reveal in this case a rather arbitrary division.

## F. FRONTAL TRANSECTS

Cluster analysis is next used to examine two frontal transects: the first from Stations 196 to 201, collected on the 6th of September 1984, and the second from Stations 283 to 291 taken 7 days later (Figure 3.1). The aim of examining these frontal transects was to investigate the different results, if any, obtained when using different combinations of attributes. Temperature, salinity and location were used in various permutations to determine which, if any, was the optimum method.

The location and mean temperature-salinity values at 10-20 m for the first transect are shown in Table VII.

Classical T-S analysis has defined the boundary between PW and AIW as the 0°C isotherm and a salinity value of 34.5. At the surface or in the near-surface layer, one can use the horizontal temperature gradient to characterise the EGPF. On this basis, one would group the stations according to:

{196, 197, 198, 199} and {200, 201}.

This is a fairly 'natural' grouping and one would expect the cluster technique to repeat this structure fairly easily. This is the case for both the simple heuristic technique and the more sophisticated iterative technique (Figure 3.16). Station 199, however, may be in a transition group rather than a definite warm or cold water station. The k-means technique, when clustering by threes, does in fact select this station as a transition regime (Figure 3.17). The simpler heuristic technique, however, identifies Station 198 with Station 199, mainly based on the close association in salinity. This is an unreasonable grouping, for as described above, temperature is the predominant property that identifies stations with respect to the EGPF.

As cluster analysis is a multi-variate statistical technique, better suited to 3 or more attributes (variables), a third parameter was selected, namely distance. In addition to temperature and salinity, a distance east or west of 5°W, the location of the EGPF, was used (Table VII). The use of distance as an additional attribute may be considered somewhat artificial. It would have been preferable to use an additional conservative water mass property, however, none was available.

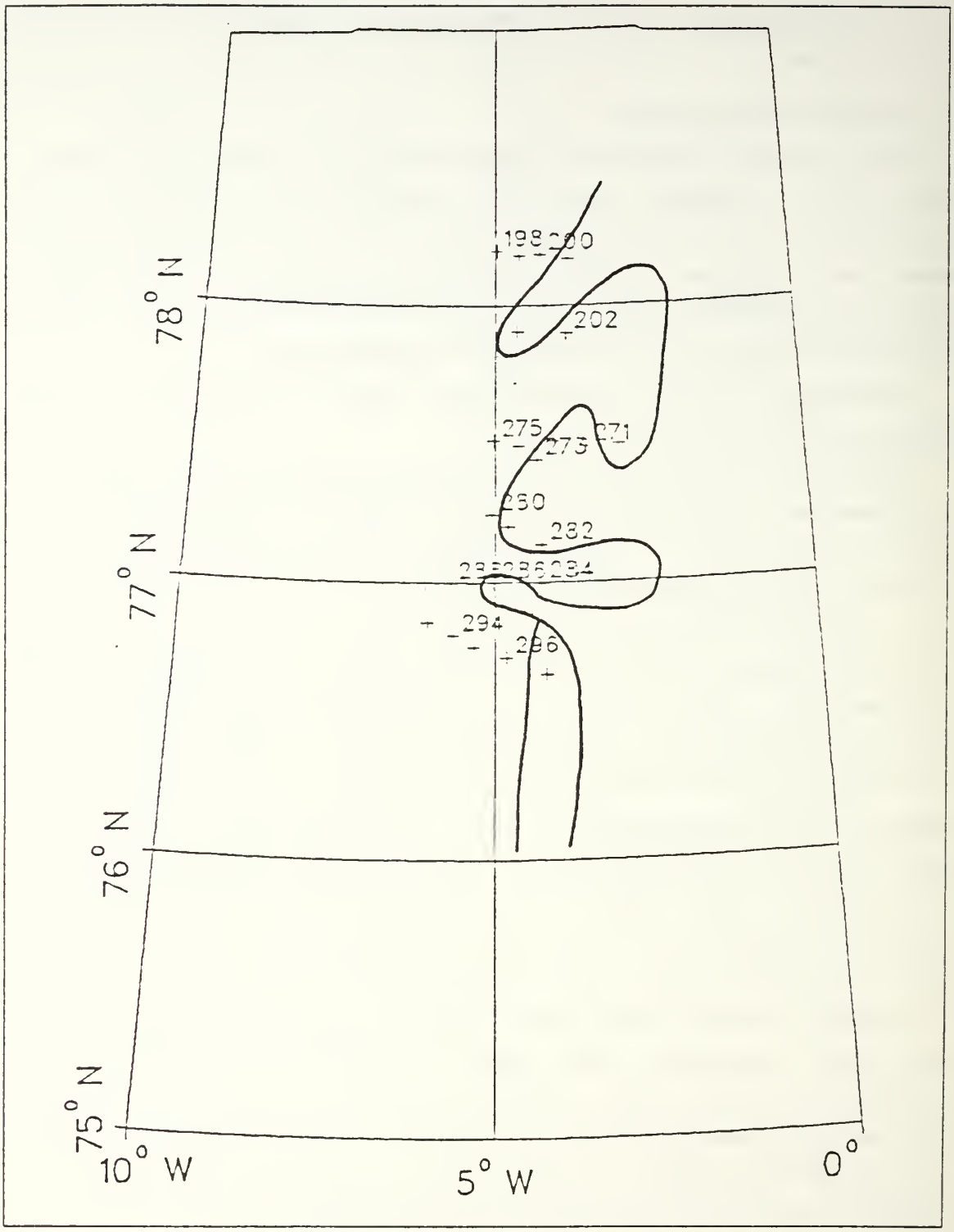


Figure 3.14 Results of a three-cluster search at 500 m for Frontal Stations using the heuristic technique.

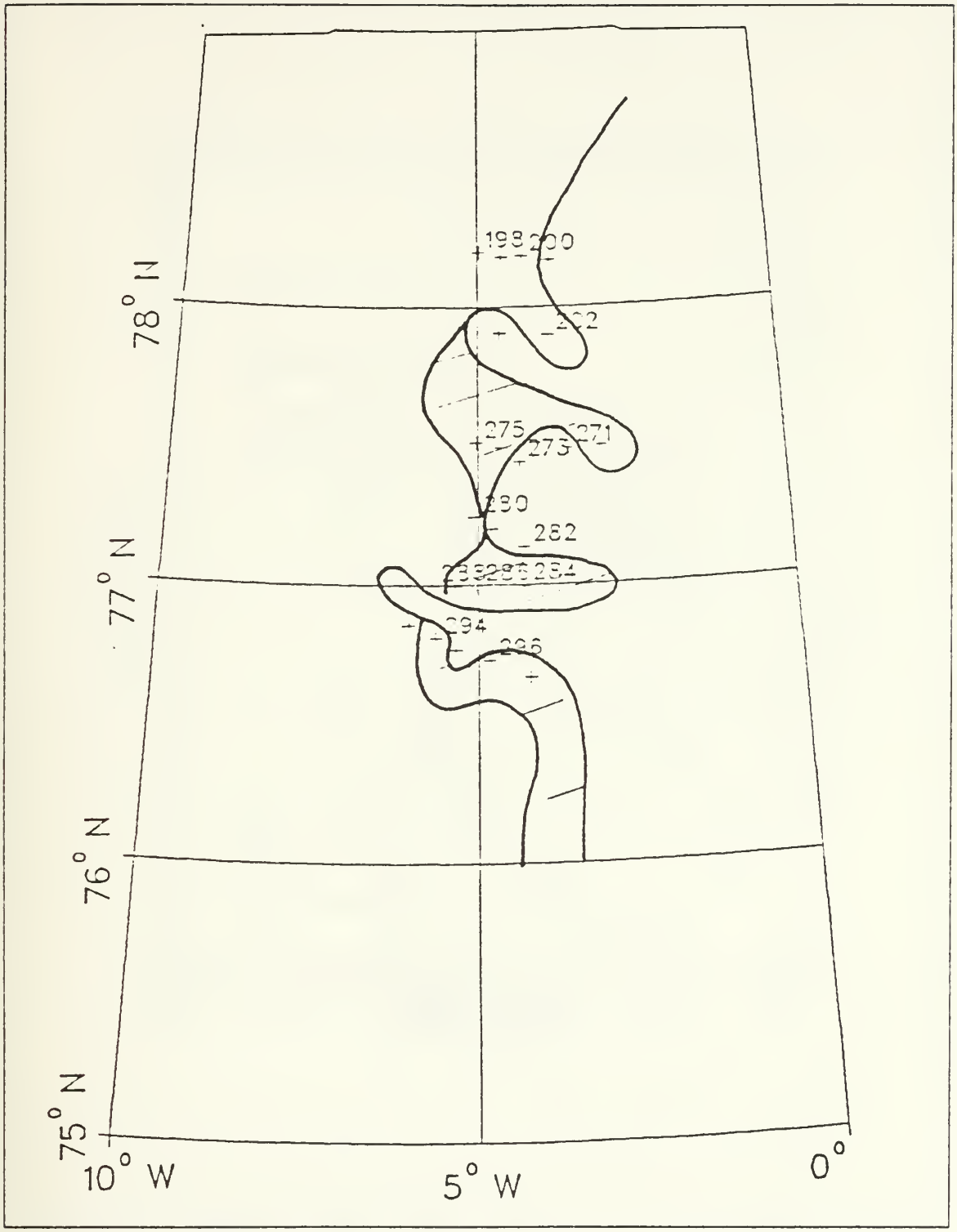


Figure 3.15 Results of a three-cluster search at 500 m for Frontal Stations using the iterative technique.

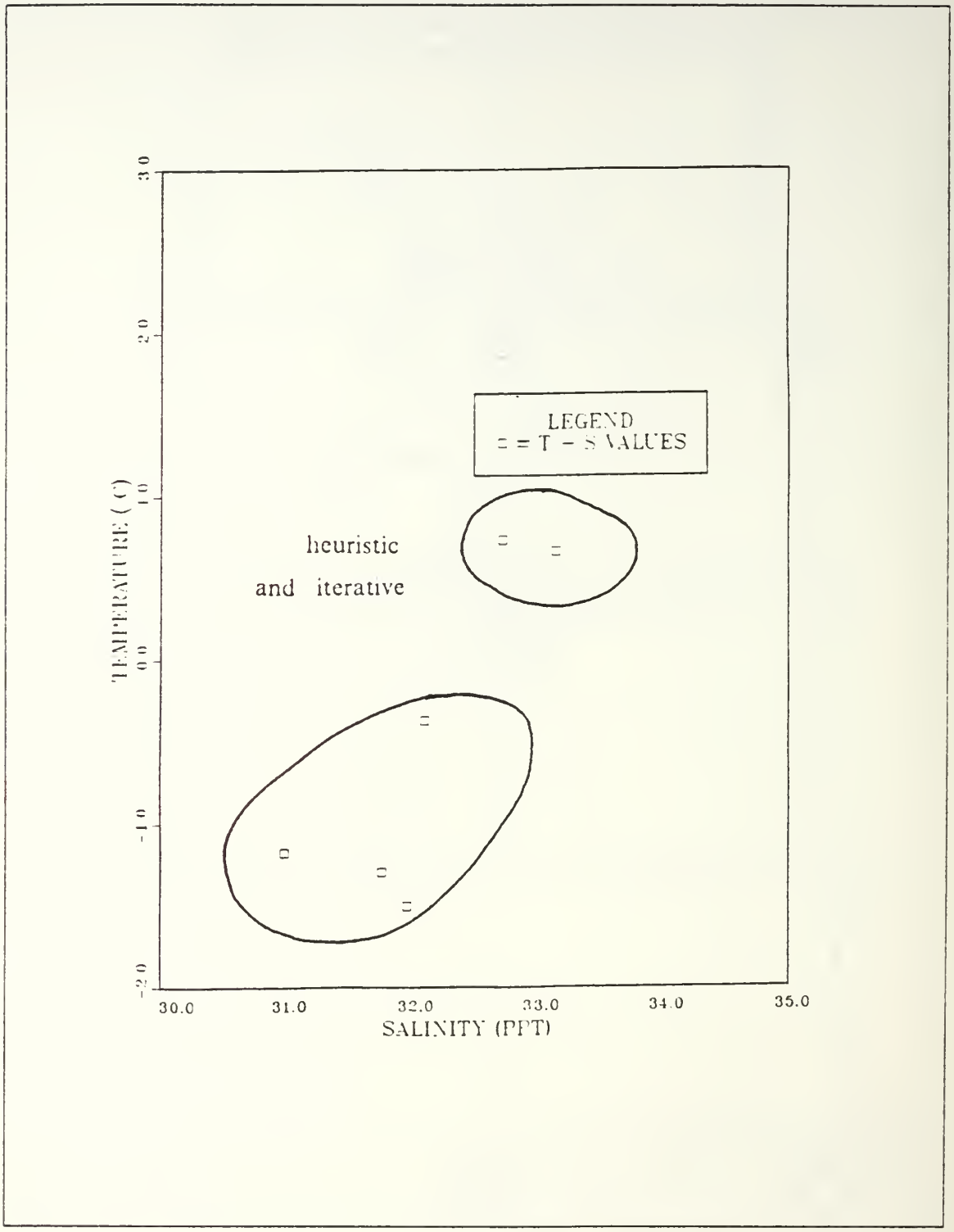


Figure 3.16 Results of a two-cluster search in the mixed layer for Stations 196-201 using the heuristic and iterative techniques.

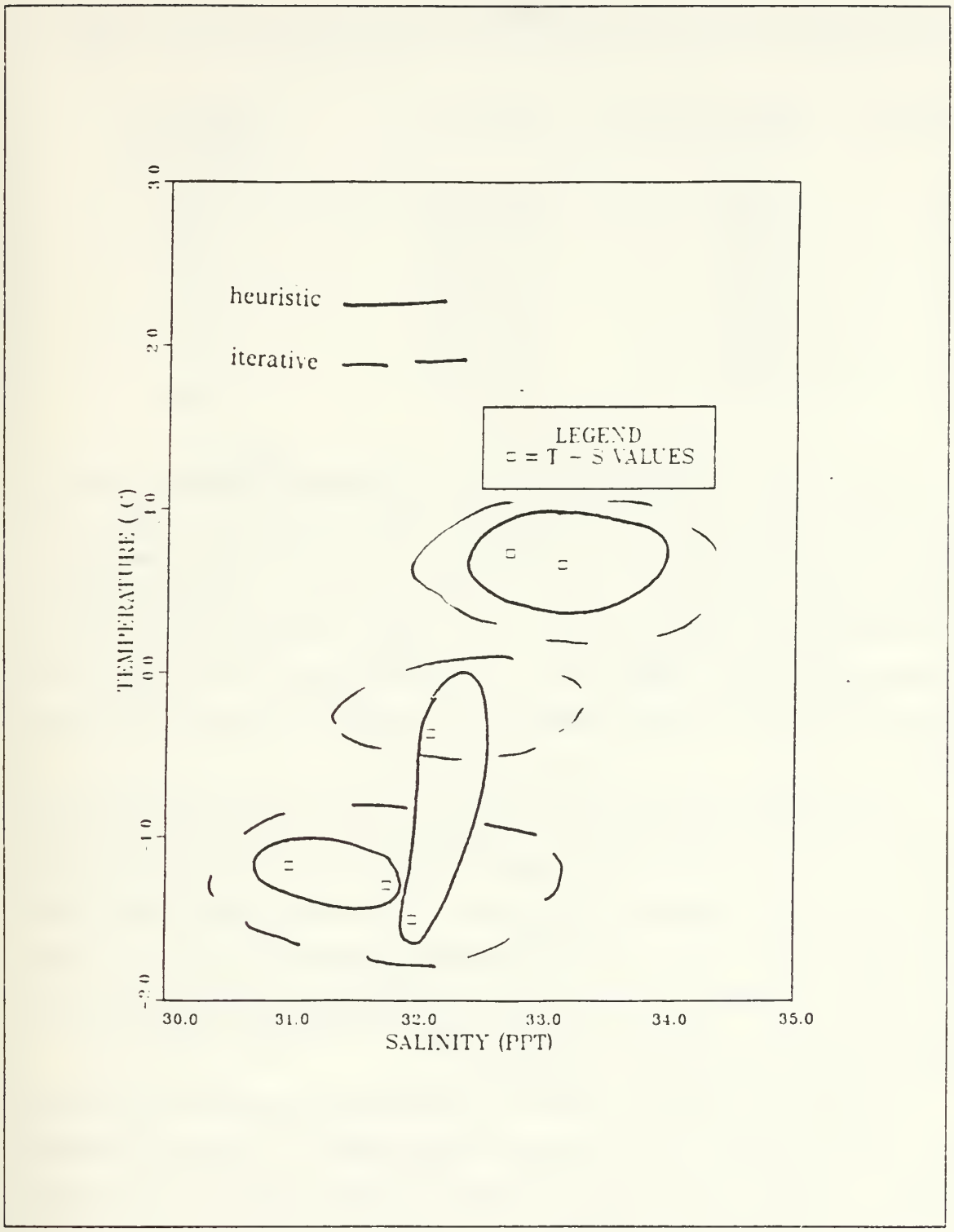


Figure 3.17 Results of a three-cluster search in the mixed layer for Stations 196-201 using the heuristic and iterative techniques.

TABLE VII

LOCATION AND MEAN T-S VALUES AT 10-20 M FOR STATIONS 196-201

| Station | Temperature(°C) | Salinity | Distance (°E or W of 5 °W) |
|---|---|---|---|
| 196 | -1.176 | 30.992 | -1.335 |
| 197 | -1.299 | 31.759 | -0.550 |
| 198 | -1.510 | 31.959 | 0.027 |
| 199 | -0.373 | 32.098 | 0.407 |
| 200 | 0.658 | 33.150 | 0.767 |
| 201 | 0.727 | 32.731 | 1.223 |

The results for a two-cluster search for both the heuristic and iterative techniques are shown in Figure 3.18. The two techniques agree in their cluster regimes which differ from the temperature-salinity case in that Station 199, which is possibly in a transition regime, is now grouped with the warm cluster regime. This result is achieved by the distance value combining with the salinity value. A similar situation occurs in the three-cluster search (Figure 3.19). Both techniques agree in their regimes and Stations 198 and 199 are paired together on the strength of their salinity and distance values.

One of the aims of this study was to determine if cluster analysis could be used as a quick, ad-hoc method of grouping data which might lead to an optimum deployment of sonobuoys, i.e., to allow for greater or lesser coherence in any one direction by using fewer or more buoys, respectively. With that in mind the next stage of the analysis considered temperature alone, as often in practice this is the only parameter that might be available. Both techniques, when using temperature alone, clustered in exactly the same fashion as when using temperature and salinity. This was the case for both a two and three-cluster search (Figures 3.16 and 3.17).

It would appear from these results, for areas where temperature is the dominant factor, that clustering by temperature alone would lead to a quick and reasonable grouping. The study now proceeds to examine the same stations at 150 m depth. Table VIII shows the values for 150 m.

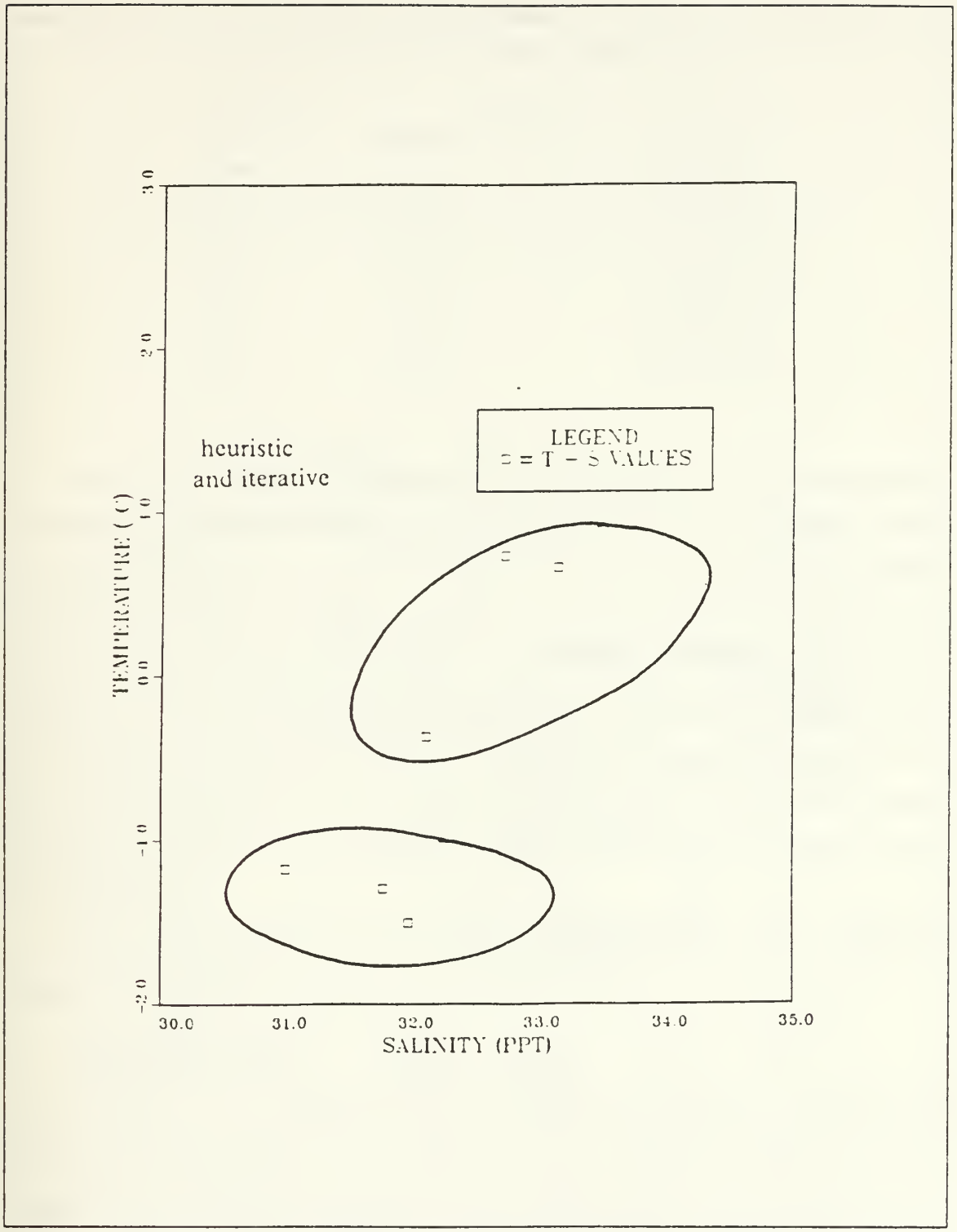


Figure 3.18 Results of a two-cluster search using three attributes for Stations 196-201.

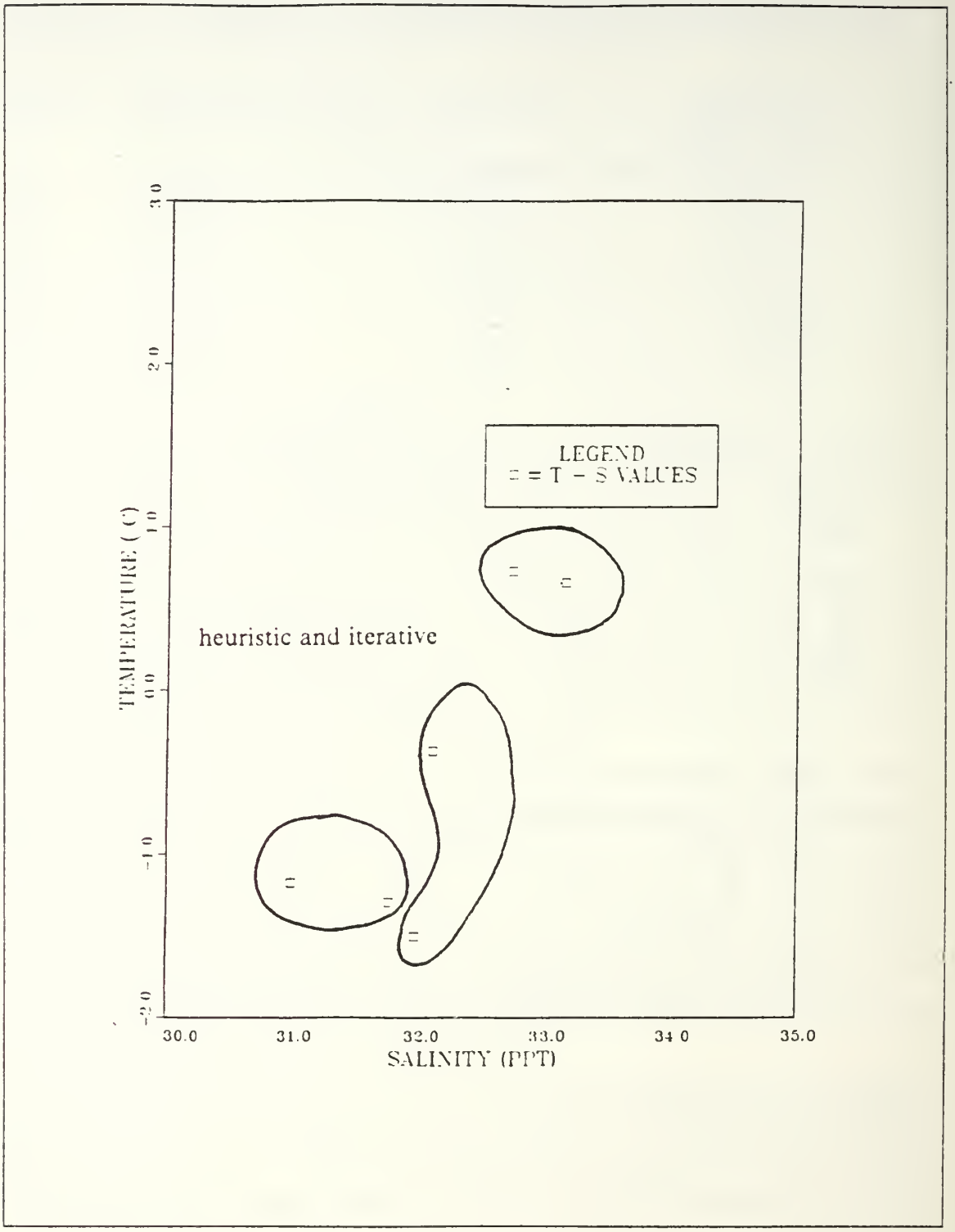


Figure 3.19 Results of a three-cluster search using three attributes for Stations 196-201.

TABLE VIII

T-S VALUES FOR STATIONS 196-201 (150 M)

| Station | Temperature(°C) | Salinity |
|---|---|---|
| 196 | -0.206 | 34.463 |
| 197 | -0.389 | 34.466 |
| 198 | 1.492 | 34.797 |
| 199 | 1.653 | 34.863 |
| 200 | 2.689 | 34.894 |
| 201 | 2.305 | 34.959 |

The salinity variations are relatively small and temperature is the distinguishing characteristic. Whether one uses the 0°C isotherm (classical) or a two-cluster analysis, the groupings are the same (Figure 3.20). There is a large temperature gradient between the two regimes and little salinity variation; hence, one would be surprised if cluster analysis did not provide this result. If three clusters are selected, both techniques yield the same result (Figure 3.21). This is an example of clustering imposing a structure. The techniques distinguish three temperature regimes; less than 0°C, less than 2°C and more than 2°C. In this instance it might be preferable to use just a two-cluster search. However, if one were dealing with a larger data set, such a result which superficially seems implausible, might reveal some subtle characteristics of the water masses.

Cluster analysis applied to the second transect yielded similar results, suggesting that the cluster technique is relatively reliable and robust and that it has a useful role to play in water mass analysis.

## G. WARM STATIONS

A further test of cluster analysis was to examine 26 stations to the east of the EGPF. The use of temperature as a single attribute, as discussed above, was used in this case. The temperature values were obtained by following a common density surface. A density (sigma-t) profile of a typical shelf station is shown in Figure 3.22. The 'knee' of the density curve has a sigma-t value of 27.8 and occurs at the depth of the maximum salinity value. Thus, the temperature corresponding to a sigma-t value

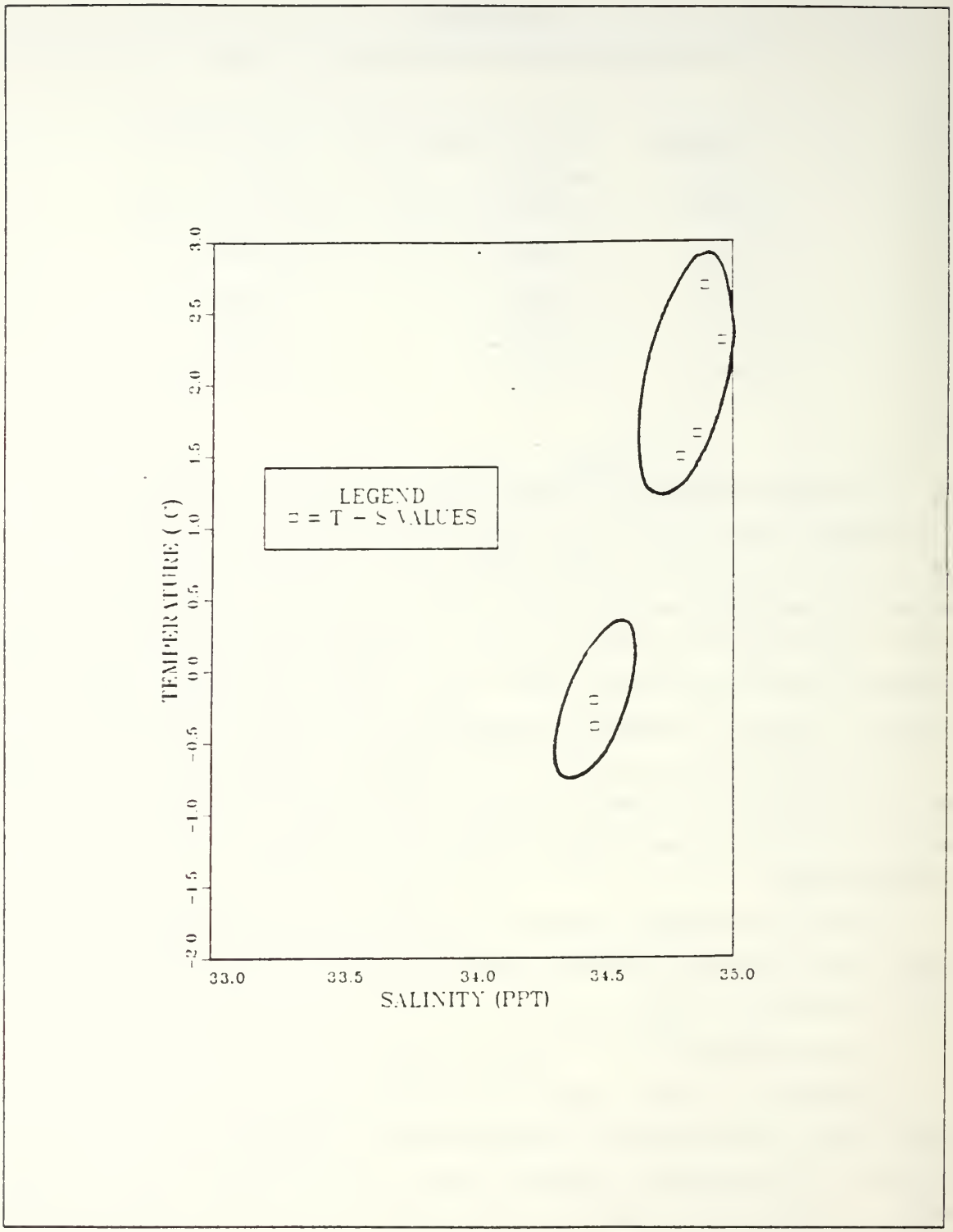


Figure 3.20 Results of a two-cluster search at 150 m for Stations 196-201 using the heuristic and iterative techniques.

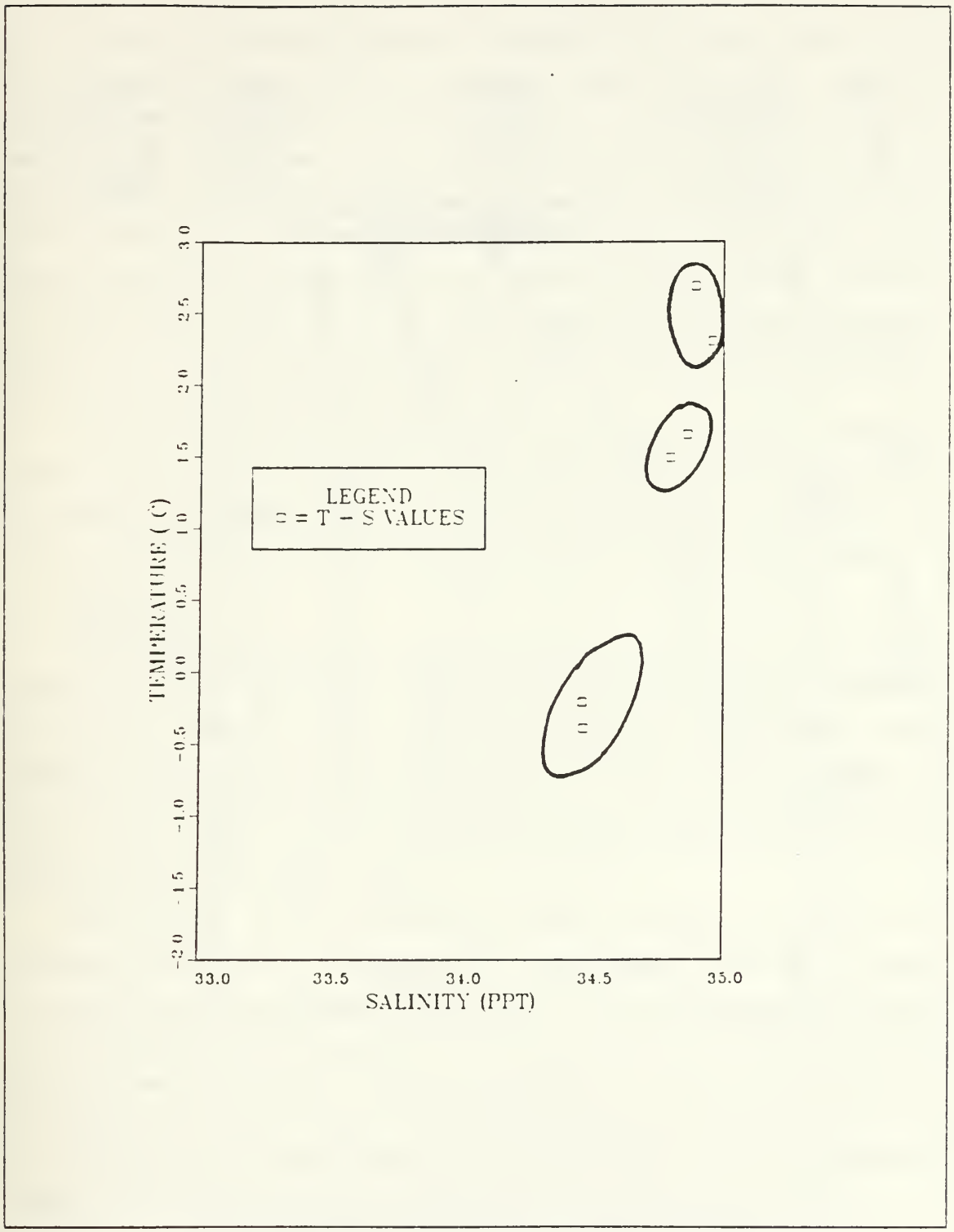


Figure 3.21 Results of a three-cluster search at 150 m for Stations 196-201 using the heuristic and iterative techniques.

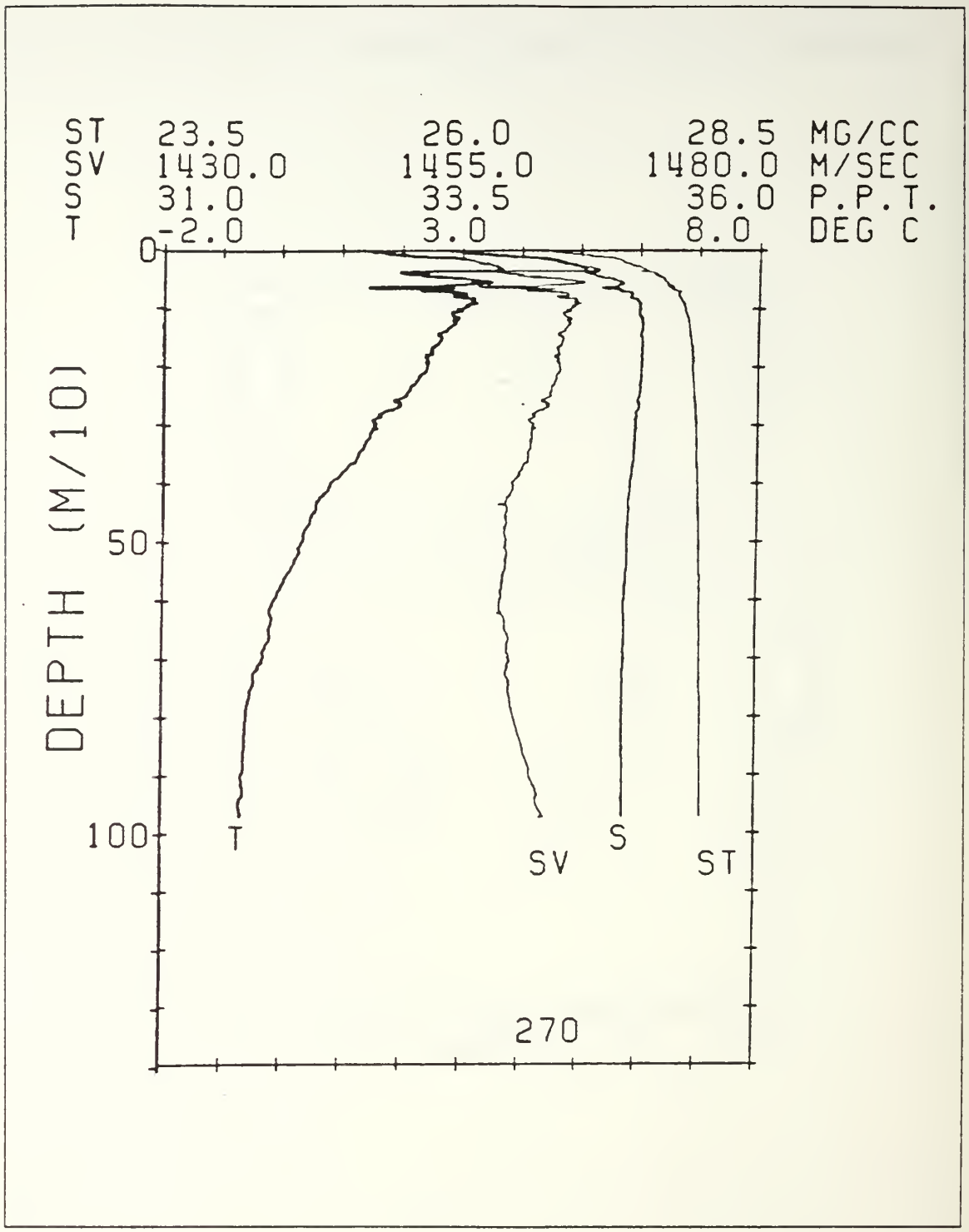


Figure 3.22 A plot of temperature, salinity, sigma-t and sound speed for a typical AIW station (Bourke and Paquette, 1985).

of 27.8 was chosen for each of the 26 stations. The data were then examined using both the heuristic and iterative techniques searching for two and three clusters. Both techniques yielded the same results for a two-cluster search (Figure 3.23). The cluster regimes comprised of the stations with 'warm' water between 77°N and 78°N, and the remainder with 'warmer' water. For a three-cluster search the two techniques are essentially similar but yield slight differences in their results. Both cluster on the basis of warm, warmer and warmest. The heuristic technique has a small group of warm stations, a large group of warmer stations and a group of only two stations comprising the warmest regime (Figure 3.24). The iterative technique picks out the cooler of the clusters of its two-cluster search as one regime and then clusters the remainder into two further regimes, warmer and warmest (Figure 3.25).

In both the two and three-cluster searches the techniques cluster in a physically realistic manner. Although this data does not contain the major frontal boundary that the previous data contained, one does see a north-south grouping with little east-west coherence. This would be useful information in planning a sonobuoy deployment.

## H. SUMMARY

The classical oceanographic method of water mass analysis, i.e., T-S analysis, identifies water masses with similar T-S properties. Cluster analysis has demonstrated that it too can repeat the 'natural' groups. In the analyses described above, cluster analysis was applied to a variety of data and in most cases identified a natural or physically meaningful grouping. The algorithm can produce clusters artificially however, as in Figures 3.14 and 3.15; with this rather unnatural grouping being distinguished by the convoluted contours.

The advantages of cluster analysis are that it is simple to apply and can accommodate many attributes simultaneously. Cluster analysis will reveal the 'shape' of the clustering, i.e., meridionally, longitudinally, circular and so on. The technique could be used operationally to ascertain the spatial coherence of data which will enable planning of sonobuoy patterns. Classical oceanographic analysis normally compares two variables (attributes) at a time. The cluster technique can easily deal with more than two attributes and offers possibilities for further research in that field. The reader is referred here to the work of Swift (1980), who uses tritium, nitrate ratios, etc., in addition to temperature and salinity to trace water masses.

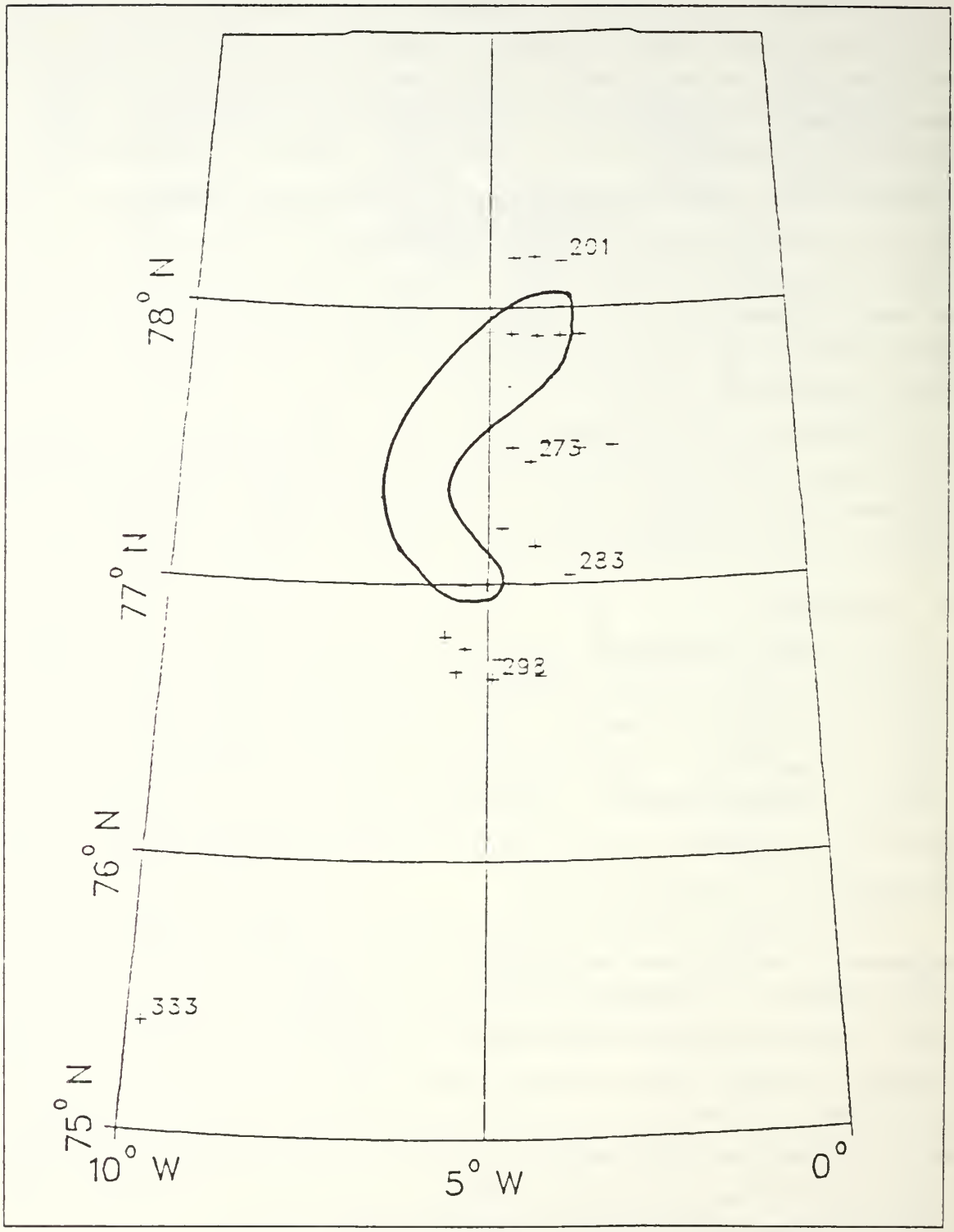


Figure 3.23 Results of a two-cluster search at a sigma-t value of 27.8 for AIW stations using the heuristic and iterative techniques.

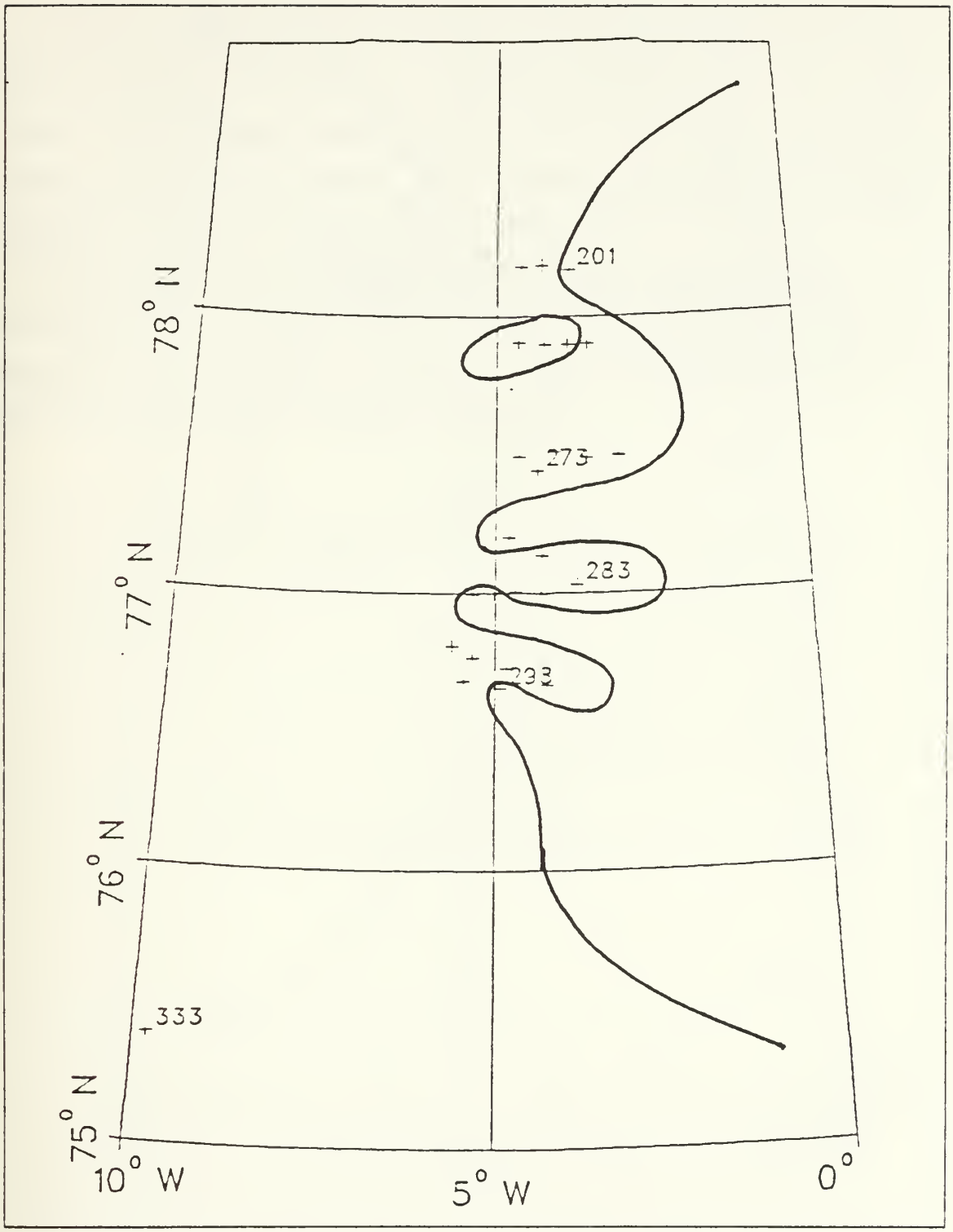


Figure 3.24 Results of a three-cluster search at a sigma-t value of 27.8 for AIW stations using the heuristic technique.

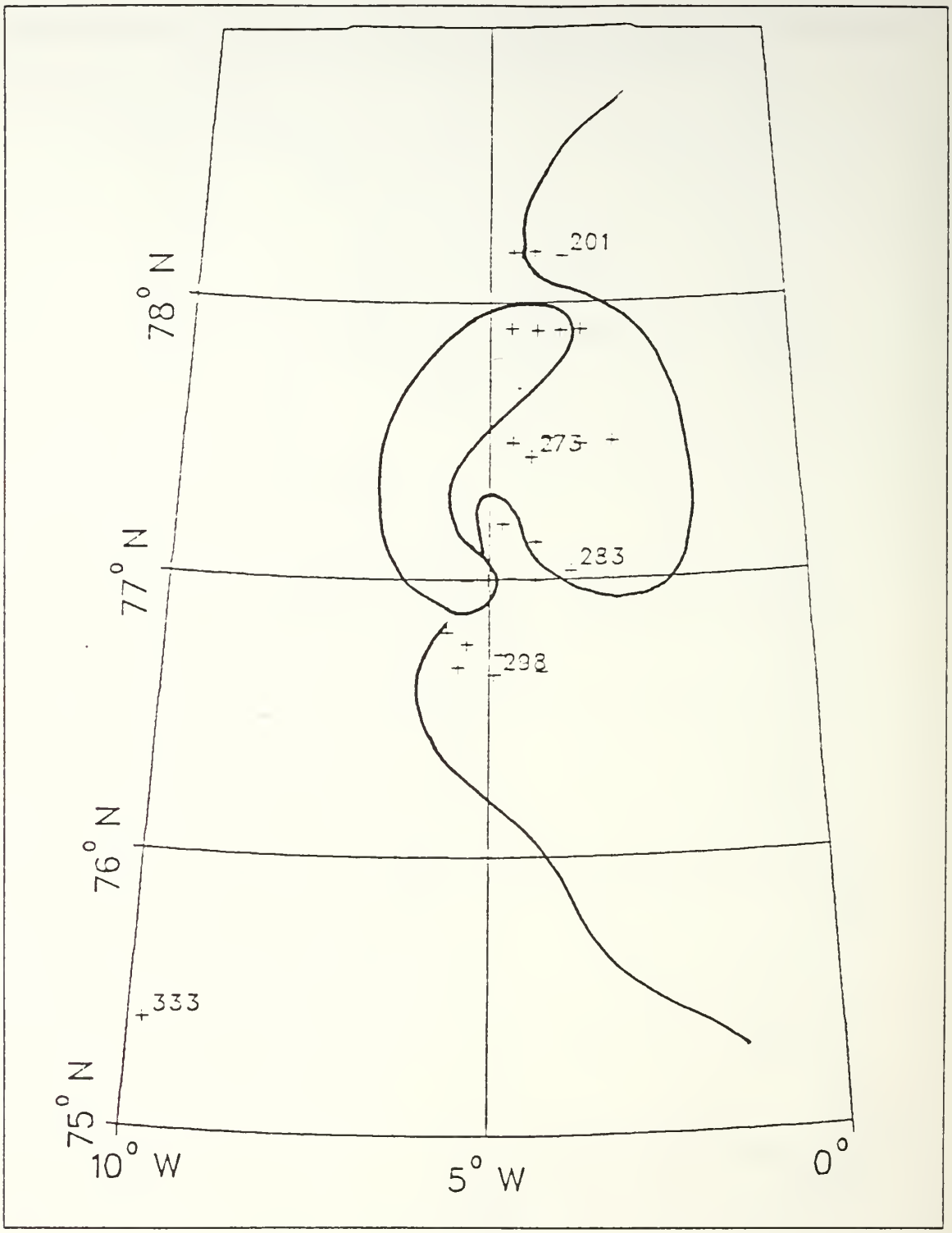


Figure 3.25 Results of a three-cluster search at a sigma-t value of 27.8 for AIW stations using the iterative technique.

The disadvantages of cluster analysis are that the technique will always find clusters (depending on RHO) even in the absence of natural groups and that some cluster techniques depend on the starting value. The latter point leads to the 'inertia' problem where the technique takes some time to resolve a new cluster. This tends to produce a cluster regime boundary that is displaced from that of the natural phenomenon. This was seen above with the heuristic technique 'displacing' the EGPF to the east of its position, as defined by classical analysis.

In conclusion, the cluster technique will not replace classical T-S analysis for characterising and identifying water masses but it does provide an efficient method for identifying the natural groups of large data sets. It also provides further potential for detailed analysis when several attributes are available.

# IV. ACOUSTICAL ANALYSIS

## A. INTRODUCTION

This chapter considers the acoustic characteristics of an area in the vicinity of the EGPF utilising available sound speed profiles (SSP). The data fall into two categories. One set is a frontal transect made perpendicularly to the front, i.e., transect A. The other is a frontal transect made obliquely to the front, transect B. The transects have one station in common and are shown in Figure 4.1.

The analysis simulates an operational situation. It is assumed that the FACT, RAYMODE and PARABOLIC EQUATION (PE) acoustic models are available. Both FACT and RAYMODE are range-independent models and hence consider only a single sound speed profile and a constant water depth throughout the propagation range. The PE model, on the other hand, can accommodate multiple sound speed profiles and varying bottom depths. Thus, the PE model is well suited in the present oceanographically complex region. The PE model will also be run in the single profile mode for comparison purposes. No measured transmission loss data were available so the PE model results, due to the more complete physics of the program, are assumed to be more valid than those from the other two models.

The next section briefly discusses the models and their input requirements. The results are then discussed. The theoretical ranges that would be forecast from the model runs are finally presented at the end of the chapter.

## B. THE ACOUSTIC MODELS

### 1. The PE Model

The PE model is a rigorous wave-theory model formulated by Brock (1978). The parabolic wave equation includes diffraction and all other full-wave affects as well as range-dependent environments. The entire range and depth-dependent acoustic field is computed as the solution is marched forward in range. The model has 35 input parameters to describe the environment and sensor dispositions. The main parameters are: bottom depth along the path, one or more sound speed profiles, source and receiver depths, half beamwidth, frequency and bottom loss data. For this study the bottom loss curves were taken from Urick (1983).

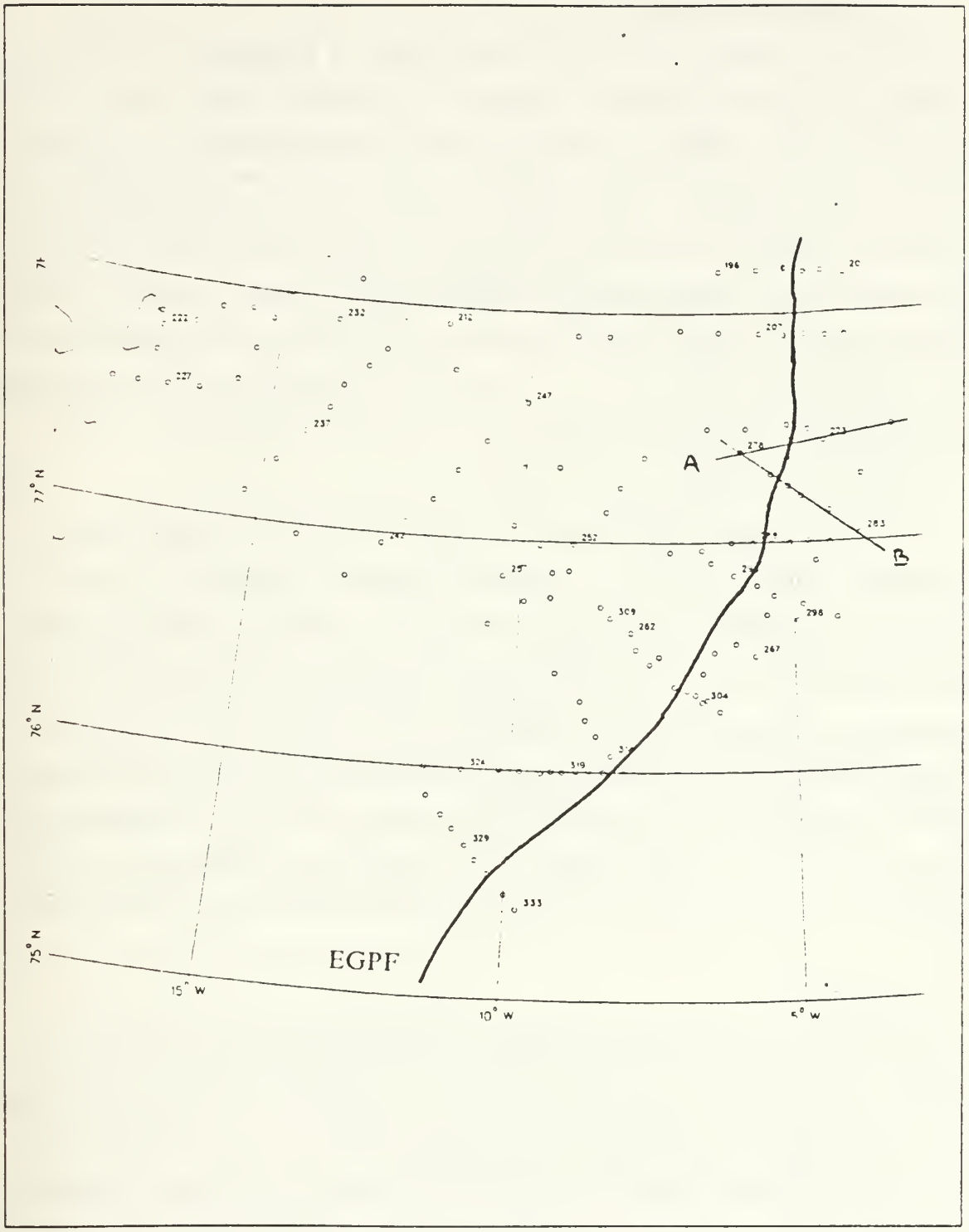


Figure 4.1 A map to show the position of the two frontal transects used for the acoustical analysis. The EGPF is also shown.

### 2. The RAYMODE Model

The RAYMODE model, as its name implies, is a ray-acoustic model and combines ray and normal mode approximations to compute acoustic energy losses (RAYMODE, 1982). This is the operational model currently used in the onboard prediction systems in the US Fleet. Ray theory is used to determine the ray bundles of interest which are classified as surface-duct, convergence zone or bottom-bounce rays. Normal mode physics are then used to compute the intensity within each ray bundle. The program uses a single sound speed profile and thus assumes a constant depth along the track. The model has a built-in family of bottom loss curves taken from the Marine Geophysical Survey bottom loss curves. Intuitively, one would expect the performance of the RAYMODE model to be worse than the PE model in a strongly range-dependent situation.

### 3. The FACT 9H Model

The Fast Asymptotic Coherent Transmission (FACT) 9H model is similar to the Raymode model in that it is a range-independent model with ray acoustics. Classical ray treatment is augmented by higher-order asymptotic corrections in the vicinity of caustics, and the phase addition of certain ray paths (Spofford, 1974). The input parameters are similar to the RAYMODE model. The major differences are in the treatment of propagation in the surface duct and the bottom loss curves. FACT calculates the intensity in the surface duct from the principle of conservation of energy modified by additional losses (proportional to range) caused by duct leakage and rough-surface scattering of energy from the duct (Marsh and Schulkin, 1967). The bottom loss is determined from the bottom loss upgrade (IBLUG) curves of Spofford (1980).

## C. ANALYSIS

The aim of this analysis is to investigate acoustic propagation across the EGPF. A variety of situations are considered. Propagation from shallow to deep water and vice versa is considered, for both the perpendicular transect A and the oblique transect B. Initially propagation is considered using a single SSP from each end of the transect. This simulates conditions when only a single SSP is available, i.e., no front is observed. In addition, propagation across the front is considered using multiple SSPs. For comparison purposes a frequency of 50 Hz is employed, although a few cases consider 300 Hz. An arbitrary figure of merit of 85 dB is used to ascertain an initial detection range (IDR).

The sound speed profiles for the three stations which provide the single profile inputs are shown in Figures 4.2, 4.3 and 4.4. Consider first the profile shown in Figure 4.2, a station in AIW to the east of the front. Although there is considerable finestructure, the essential acoustic features are relatively simple. There is a strong positive gradient from the surface to 35 m. Below this the finestructure creates two narrow sound channels from 35 m to 55 m and from 55 m to 75 m. This is mainly caused by the interleaving of warm and cold water, from either side of the EGPF. Below 85 m a weak negative gradient extends to the local minimum of 1458 m/s at 560 m. Thereafter the sound speed gradient increases slowly, giving a relatively weak channel between 175 m and the ocean bottom.

The profile of a typical shelf station is shown in Figure 4.3. The water depth is shallow, 250 m, and the sound speed profile mirrors the temperature profile. The profile is slightly positive from the surface to 20 m. Below 20 m a weak channel exists between 20 m and 125 m, with its axis at 65 m, the depth of the coldest water. At 125 m the water mass changes rapidly from PW to AIW and the corresponding temperature increase is reflected in the increase in sound speed, changing 13 m/sec over 75 m. Over the last 50 m there is little change in temperature or salinity but the slight positive gradient in sound speed indicates the influence of pressure.

The third profile is shown in Figure 4.4. This profile is located in AIW to the east of the front, and is similar to that shown in Figure 4.2, without the finestructure of the upper 100 m. This profile shows a positive gradient of 10 m/sec over the top 65 m, which is less than the 12 m/sec in the first 35 m shown in Figure 4.2. There is a weak negative gradient to 455 m and then a weak positive gradient extending to the bottom. Thus a strong duct is present in the upper 65 m and a sound channel from 75 m or so to the ocean bottom, with its axis at 455 m.

The three sound speed profiles discussed above simulate the information available operationally. It would be preferable to use multiple SSPs utilising all the information from source to receiver, especially when crossing such a significant feature as the EGPF. To this end, all the the sound speed profiles of the frontal transects are considered by using them as inputs to the PE model. The multiple profiles are shown in Figures 4.5 and 4.6. Figure 4.5 shows the profiles of station 270 to 277, the perpendicular transect. Figure 4.6 shows the profiles of stations 277 to 283, the oblique transect. The PE model is run for both east to west and west to east simulations.

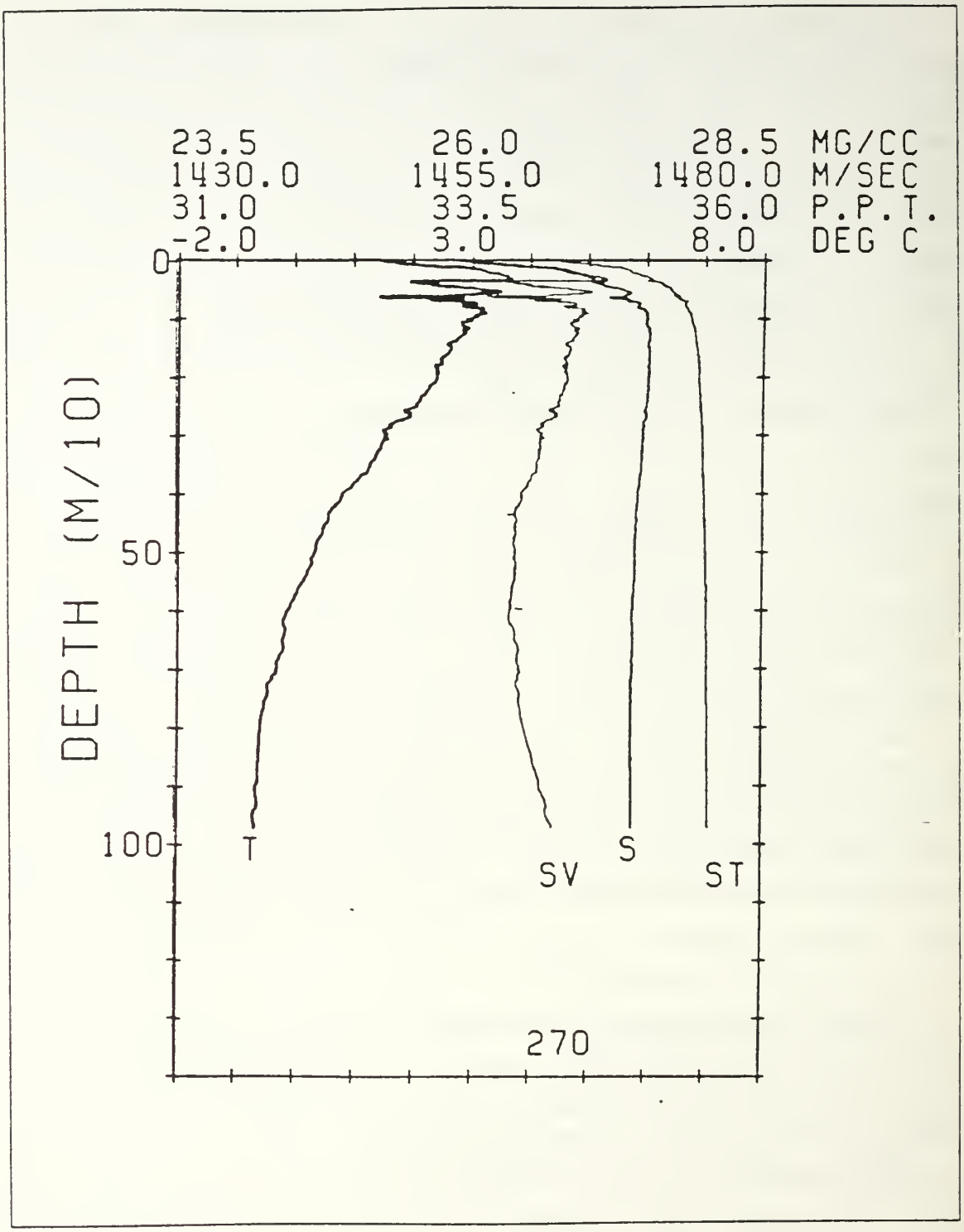


Figure 4.2 A property plot of Station 270, in deep AIW to the east of the EGPF (Bourke and Paquette, 1985).

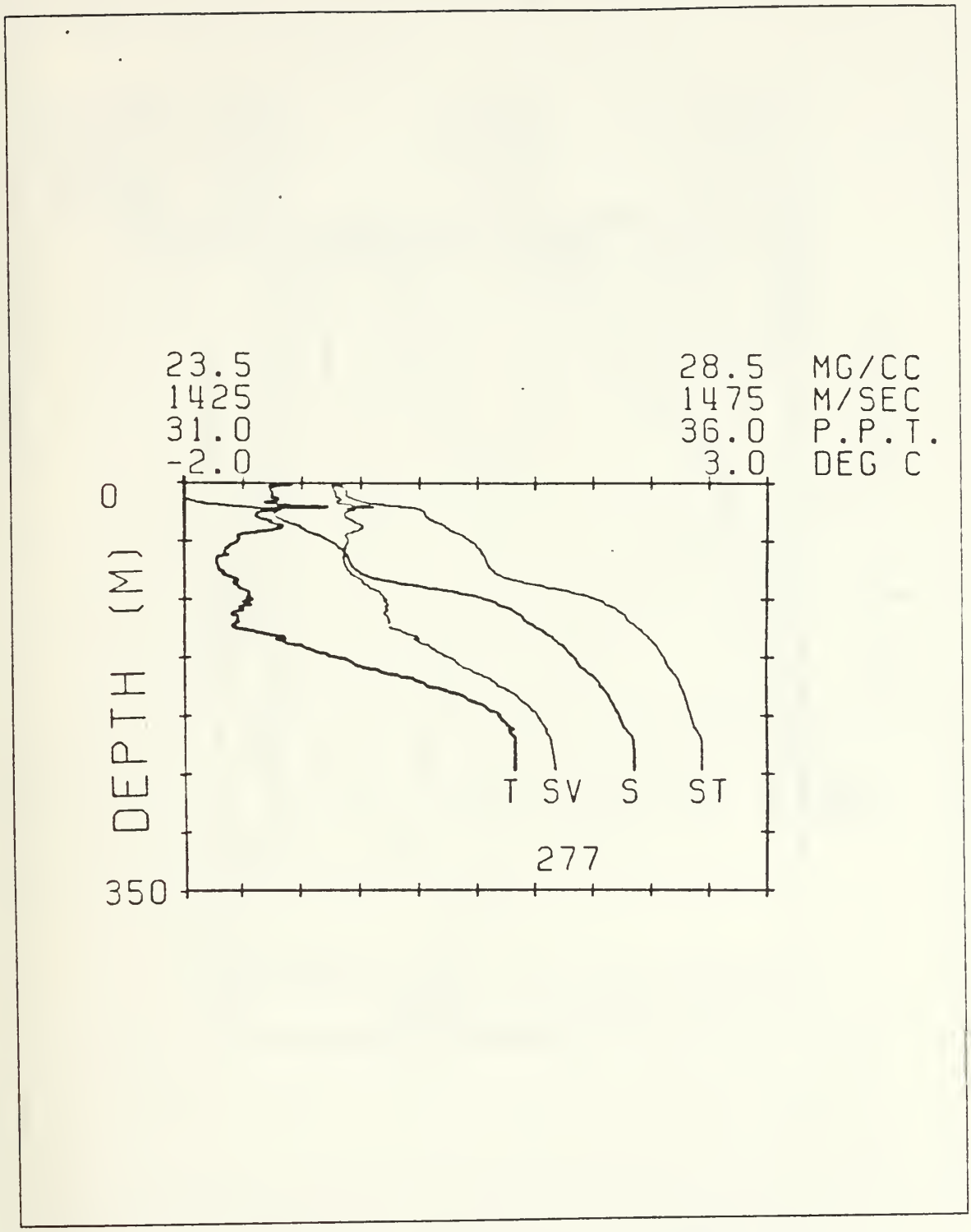


Figure 4.3 A property plot of Station 277, in shallow PW on the continental shelf (Bourke and Paquette, 1985).

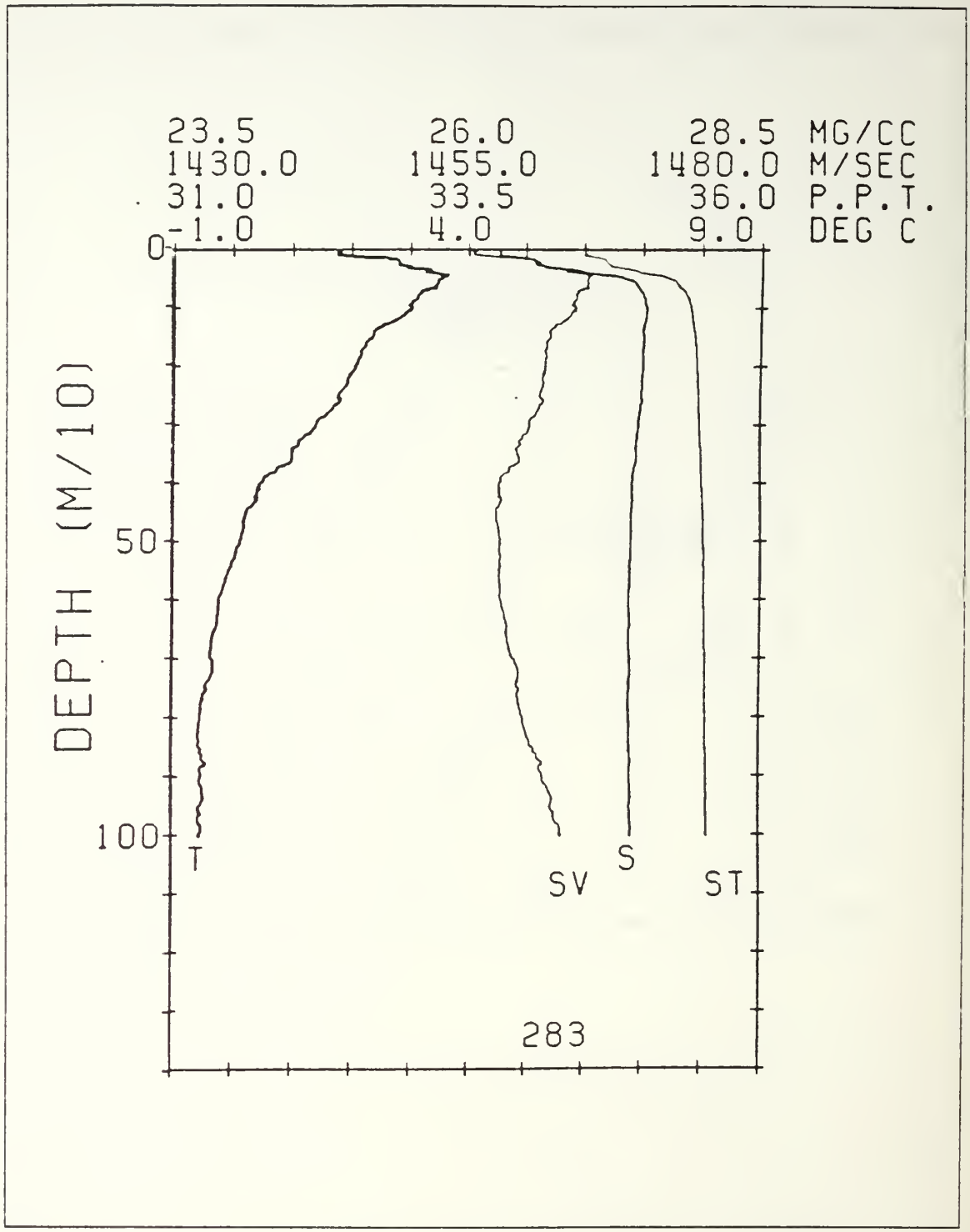


Figure 4.4 A property plot of Station 283, in deep AIW to the east of the EGPF (Bourke and Paquette, 1985).

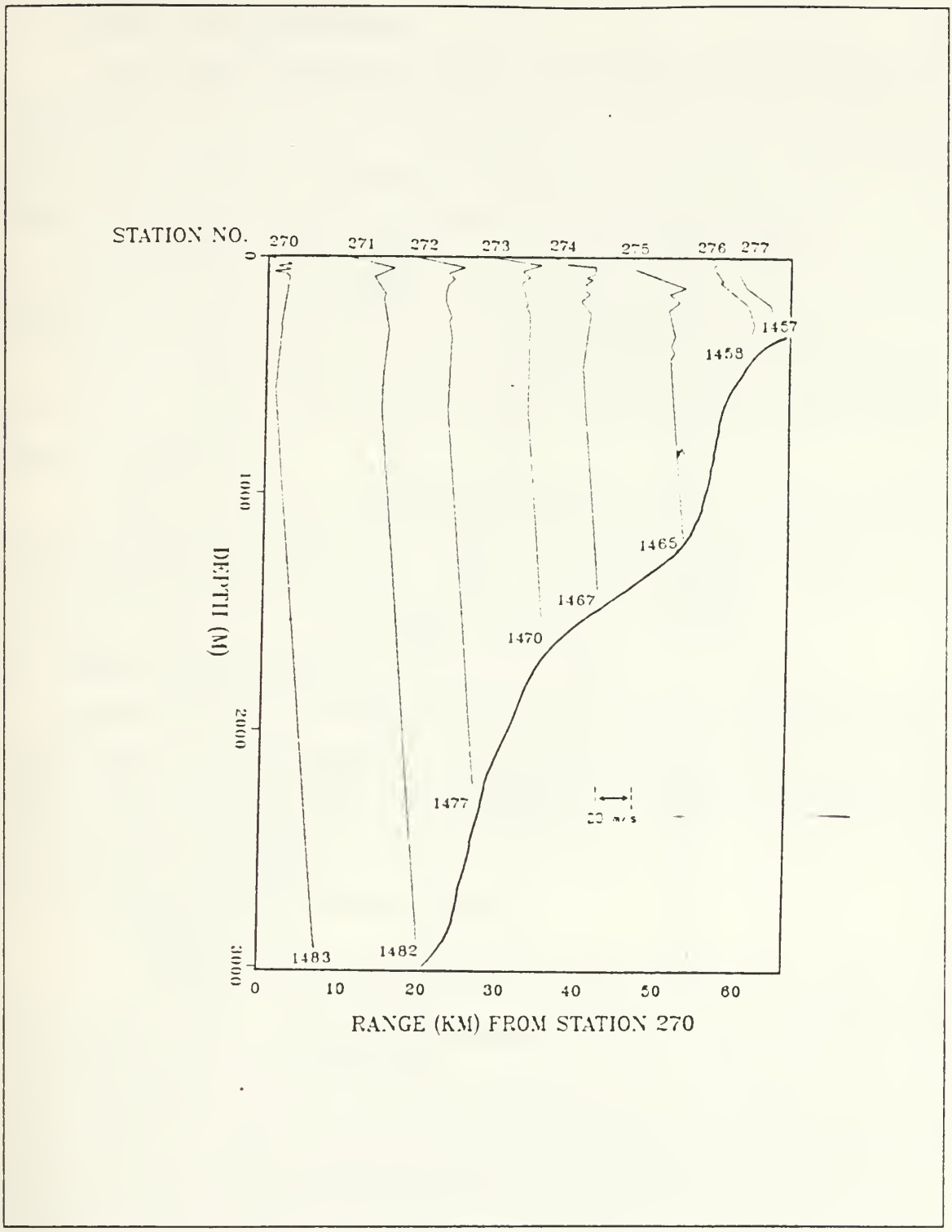


Figure 4.5 A sound speed plot of Stations 270 - 277, a perpendicular transect. Sound speed at the bottom of the water column is indicated.

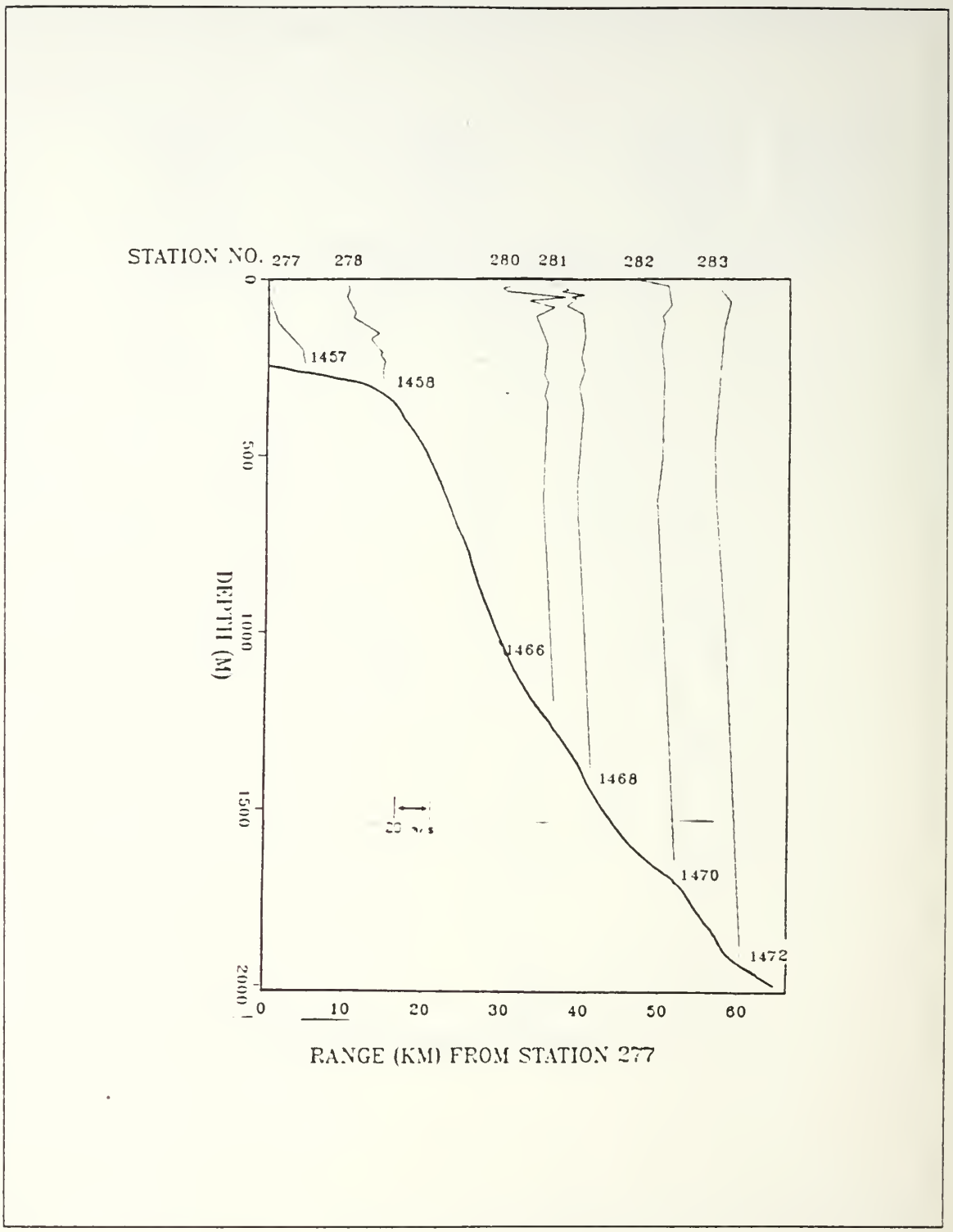


Figure 4.6 A sound speed plot of Stations 277 - 283, an oblique transect. Sound speed at the bottom of the water column is indicated.

1. **Source - Receiver Dispositions**

For the single profile shown in Figure 4.2 (to the east of the front), the source is placed at 20 m. The receiver is placed first at 65 m, on the axis of one of the weak sound channels, then also at 150 m. The latter position is in the weak negative gradient in the upper half of the SOFAR channel. In addition, both the source and receiver are placed at 65 m to test a completely channelled propagation path.

For the shelf station (Figure 4.3), the source is placed at 18 m in the surface duct. The receiver is placed at 65 m or on the axis of the sound channel. In the case of the deep water environment, both source and receiver are placed at 65 m.

For the more southerly of the two warm water stations (Figure 4.4) the cross layer case of a source in the surface duct with a receiver at 150 m in the SOFAR channel is considered.

When all SSPs are considered for the PE model runs, the acoustic environment from east to west and from west to east is examined separately for each of the two frontal transects. This is done to test for acoustic reciprocity. Similar source and receiver combinations to those described above are considered.

## D. TRANSMISSION LOSS

Propagation loss curves were generated to obtain the predicted ranges; for clarity only the forecast ranges are presented. These ranges are listed in Tables IX to XVI inclusive. This section considers the transmission loss as it affects forecast ranges and the final discussion section considers the differences between the models, path orientation and acoustic reciprocity.

1. **From Deep Water (looking shoreward)**

*a. Single Profiles*

Tables IX to XI present the ranges that would be forecast using the different models. For a source at 20 m and a receiver at 65 m, FACT forecasts 11 km, RAYMODE over 65 km and PE 5 km. A similar situation occurs with a source - receiver geometry of 20/150 m with FACT predicting 14 km, RAYMODE over 65 km and PE 6 km. It is evident that the FACT range is over two-hundred per cent greater than the PE range and RAYMODE forecasts extended ranges that are more than a thousand times greater than PE. RAYMODE continues to be optimistic with the same source - receiver depths at 300 Hz, predicting 31 km compared with a FACT range of 13 km. The only agreement shown in the tables is that between FACT and PE when

both source and receiver are in the sound channel at 65 m (for 300 Hz) when ranges are in excess of 65 km.

TABLE IX
PREDICTED RANGES (FACT)

Deep water, Single Profile

| Source (m) | Receiver (m) | Frequency (Hz) | IDR (km) |
|---|---|---|---|
| 20 | 65 | 50 | 11 |
| 20 | 150 | 50 | 14 |
| 65 | 65 | 300 | over 65 |
| 20 | 150 | 300 | 13 |

TABLE X
PREDICTED RANGES (RAYMODE)

Deep water, Single Profile

| Source (m) | Receiver (m) | Frequency (Hz) | IDR (km) |
|---|---|---|---|
| 20 | 65 | 50 | over 65 |
| 20 | 150 | 50 | over 65 |
| 20 | 150 | 300 | 31 |

*b. Multiple Profiles*

Tables XII and XIII present the ranges forecast by the PE model for a perpendicular and an oblique transect. Propagation perpendicular to the EGPF is greater in both the 150/60 m and 150/150 m source - receiver combinations. In the first case the perpendicular range is 9 km compared with an oblique range of 5 km. This means that the oblique range is only 55% of the perpendicular range. Similarly, at 150/150 m the oblique range is 64% of the perpendicular range.

TABLE XI
PREDICTED RANGES (PE)

Deep water, Single Profile

| Source (m) | Receiver (m) | Frequency (Hz) | IDR (km) |
|---|---|---|---|
| 20 | 65 | 50 | 5 |
| 20 | 150 | 50 | 6 |
| 65 | 65 | 300 | over 65 |
| 60 | 60 | 50 | 7* |

* sloping bottom

TABLE XII
PREDICTED RANGES (PE)

East to west, Multi Profile
Perpendicular Transect

| Source (m) | Receiver (m) | Frequency (Hz) | IDR (km) |
|---|---|---|---|
| 20 | 15 | 50 | 2 |
| 60 | 60 | 50 | 8 |
| 150 | 15 | 50 | 6 |
| 150 | 60 | 50 | 9 |
| 150 | 150 | 50 | 11 |
| 20 | 65 | 50 | 6 |
| 20 | 150 | 50 | 7 |

2. **From Shallow Water (looking east)**

*a. Single Profiles*

Tables XIV to XVI show that all 3 models predict detection ranges to the end of the transects and beyond for all source - receiver geometries. Clearly the strongly ducted environment permits significant trapping of acoustic energy, thus ensuring that all ranges extend beyond the 70 km limit of each plot.

TABLE XIII

PREDICTED RANGES (PE)

East to west, Multi Profile

Oblique Transect

| Source (m) | Receiver (m) | Frequency (Hz) | IDR (km) |
|---|---|---|---|
| 15 | 15 | 50 | 2 |
| 60 | 15 | 50 | 3 |
| 150 | 60 | 50 | 5 |
| 150 | 150 | 50 | 7 |
| 20 | 65 | 50 | 3 |
| 20 | 150 | 50 | 4 |

TABLE XIV

PREDICTED RANGES (FACT)

West to east, Single Profile

| Source (m) | Receiver (m) | Frequency (Hz) | IDR (km) |
|---|---|---|---|
| 15 | 60 | 50 | over 65 |
| 15 | 150 | 50 | 65 |
| 15 | 150 | 300 | 65 |
| 60 | 60 | 300 | over 65 |

TABLE XV

PREDICTED RANGES (RAYMODE)

West to east, Single Profile

| Source (m) | Receiver (m) | Frequency (Hz) | IDR (km) |
|---|---|---|---|
| 15 | 60 | 50 | over 65 |
| 15 | 150 | 50 | over 65 |
| 15 | 150 | 300 | 65 |

TABLE XVI
PREDICTED RANGES (PE)

West to east, Single Profile

| Source (m) | Receiver (m) | Frequency (Hz) | IDR (km) |
|---|---|---|---|
| 15 | 60 | 50 | over 65 |
| 15 | 150 | 50 | over 65 |
| 60 | 60 | 300 | over 65 |

*b. Multiple Profiles*

The ranges for the multiple SSPs are longer than the east to west case, both for oblique and normal propagation (Tables XVII and XVIII). Comparing the 15/60 m source - receiver depths, a range of 41 km is forecast for a perpendicular transect. The oblique transect forecasts 35 km, about 85% of the previous case. For a source - receiver disposition of 15/150 m, the oblique transect forecasts 77% of the range of the perpendicular transect. For a source - receiver depth of 150/150 m the ranges are 65 km for perpendicular propagation and 54 km for oblique, 83% of the former. Similarly, with deep to shallow water propagation, there is a significant reduction of range for the oblique propagation case compared with the orthogonal case.

## E. DISCUSSION

The purpose of this chapter was to investigate the differences, if any, in range prediction between models, in propagation both normal and oblique to the EGPF and also reciprocity. These results are discussed below.

### 1. Model Differences

The range-dependent PE model is able to incorporate multiple SSPs and bottom slopes and is assumed to be the "best" model, i.e., the most suitable of the three models considered. Thus, the results of the PE model are used as the basis for comparison. For a west to east comparison, all three models using single SSPs gave the same results, i.e., an initial detection range of over 65 km. The PE model, when used with multiple SSPs, and a source-receiver combination of 15/60 m, forecasts ranges of 41 km and 35 km for the orthogonal and oblique case, respectively. The situation is different for a source - receiver geometry of 15/150 m, where the multiple SSP PE forecast of 65 km is in agreement with the single profile models.

TABLE XVII
PREDICTED RANGES (PE)

West to east, Multi Profile
Perpendicular Transect

| Source (m) | Receiver (m) | Frequency (Hz) | IDR (km) |
|---|---|---|---|
| 15 | 60 | 50 | 41 |
| 15 | 150 | 50 | 65 |
| 60 | 60 | 50 | over 65 |
| 100 | 60 | 50 | 63 |
| 150 | 150 | 50 | 65 |
| 60 | 60 | 50 | 9* |

* constant water depth

TABLE XVIII
PREDICTED RANGES (PE)

West to east, Multi Profile
Oblique Transect

| Source (m) | Receiver (m) | Frequency (Hz) | IDR (km) |
|---|---|---|---|
| 15 | 60 | 50 | 35 |
| 15 | 150 | 50 | 50 |
| 60 | 150 | 50 | 37 |
| 150 | 150 | 50 | 54 |

The strong positive gradient found in the shallow, continental shelf waters would be expected to produce relatively long ranges, as much of the energy is waterborne, due to the strong focusing in the surface duct. This focusing of energy produces the enhanced ranges of the single profile models. The multiple-profile PE model is also heavily influenced by this initially strong focusing. In addition, the presence of the EGPF is diminished as it too contains a duct and energy remains trapped in it. The reduced ranges for the 15/60 m case are due to the receiver SSP profile. The receiver depth of 60 m is a local sound speed maximum leading to ray

divergence and this reduces the forecast range when compared with the 15/150 m geometry.

The east to west, or shoreward looking ranges highlight significant differences between the models. Single SSP ranges are similar to those from multiple SSPs, underestimating detection ranges by about 15%. The enhanced range of the PE model is probably due to a focusing of acoustic energy as the shallow water profiles are incorporated into the multiple SSP model, thus increasing the ranges. The FACT model is rather optimistic in that its ranges are about twice those of the multiple SSP PE model. RAYMODE, however, is unreliable with predicted ranges in excess of 65 km, some ten times longer than the assumed best answer. The cases considered for this shoreward-looking case have the source in the surface duct at 20 m with the receiver below the duct at 65 m and 150 m. FACT deals with this cross-duct case by reducing intensity by 10 dB, which would appear to be slightly optimistic. RAYMODE, in contrast, deals very poorly with this case seemingly regarding the whole profile as an extended duct and producing the excessive ranges referred to above.

2. **Normal and Oblique Transects**

While oblique ranges are less than those normal to the front, there is a further distinction between west to east and east to west propagation. For east to west propagation the oblique ranges varied from 55 to 64% of the perpendicular ranges for various source/receiver combinations, whereas in contrast, the west to east oblique ranges were 77 to 85% of the orthogonal ones. The greater similarity of the latter case is probably due to the fact that both transects share the same source profile. This common profile is from shallow water with the strong positive gradient trapping significant amounts of energy. It is only the receiver profiles which differ and it is the source profile which has more influence on propagation. In contrast, for east to west propagation the orthogonal and oblique cases have different source profiles. As has been pointed out above, the source profile for the perpendicular case has a positive gradient of 12 m/sec in the upper 35 m, whereas for the oblique case the gradient change is 10 m/sec in 65 m. In addition, there is an absence of finestructure in the source profile for the oblique transect. Thus, it is probable that source or receiver profile differences are the major factor in producing the different ranges, although this area is one which would benefit from further study.

3. **Acoustic Reciprocity**

As described above to test for acoustic reciprocity both east to west and west to east profiles were input to the PE model. Ranges for a deep water source to a shallow water receiver are shown in Table XII, those for the reverse direction in Table XVII. The two source - receiver geometries chosen for comparison were 60/60 m and 150/150 m. The 60/60 m case has a predicted range of 8 km when considered from deep to shallow, whereas the range is over 65 km for the reverse. The difference in the 150/150 m case is of the same order.

The propagation loss (PL) curves for the 60/60 m case are shown in Figures 4.7 to 4.10. The PL curve for a deep water receiver (Figure 4.7) shows a rapid fall off of energy to a minimum of 96 dB at 12 km. There is a broad CZ region between 20 and 28 km (for a FOM of 85 dB) with a second CZ between 40 and 55 km. The signal excess is a maximum of 6 dB and 12 dB in the first and second CZs, respectively.

In contrast, the PL curve for the shallow water source to deep water receiver (Figure 4.8) shows the influence of the shallow water profile. The highly positive gradient and consequent trapping of energy results in a significantly different shape of the PL curve. A signal excess of at least 15 dB to 20 km is noted; beyond this range there is a significant increase in transmission loss. However, there remains a mean signal excess of some 3 - 5 dB to 65 km.

In an attempt to distinguish between multiple profile and sloping bottom affects, two further PL curves are considered (Figures 4.9 and 4.10). The case of the multiple profile with a flat bottom is shown in Figure 4.9. This curve is virtually identical to the upward sloping bottom curve shown in Figure 4.7. Again this emphasises the importance of the source profile in determining acoustic propagation. This was also seen in the relative similarity of ranges when using either a single or multiple SSP PE model. The range predicted using this multiple SSP, flat-bottomed curve would be 9 km compared with 8 km for the assumed best answer. A different picture emerges when using a single SSP but an upward sloping bottom, Figure 4.10. The single SSP used was the deep water profile at the eastern end of the orthogonal transect. Initially, the PL curve looks similar, in that a range of 7 km would be predicted, i.e., transmission loss is 85 dB at 7 km. However, the signal excess falls to a minimum of -40 dB at 20 km, which the multiple SSP PL curves indicate as the range to the first CZ. The influence of the deep water profile is paramount over this first 20 km with energy quickly spreading out and dissipating. Gradually the upward sloping

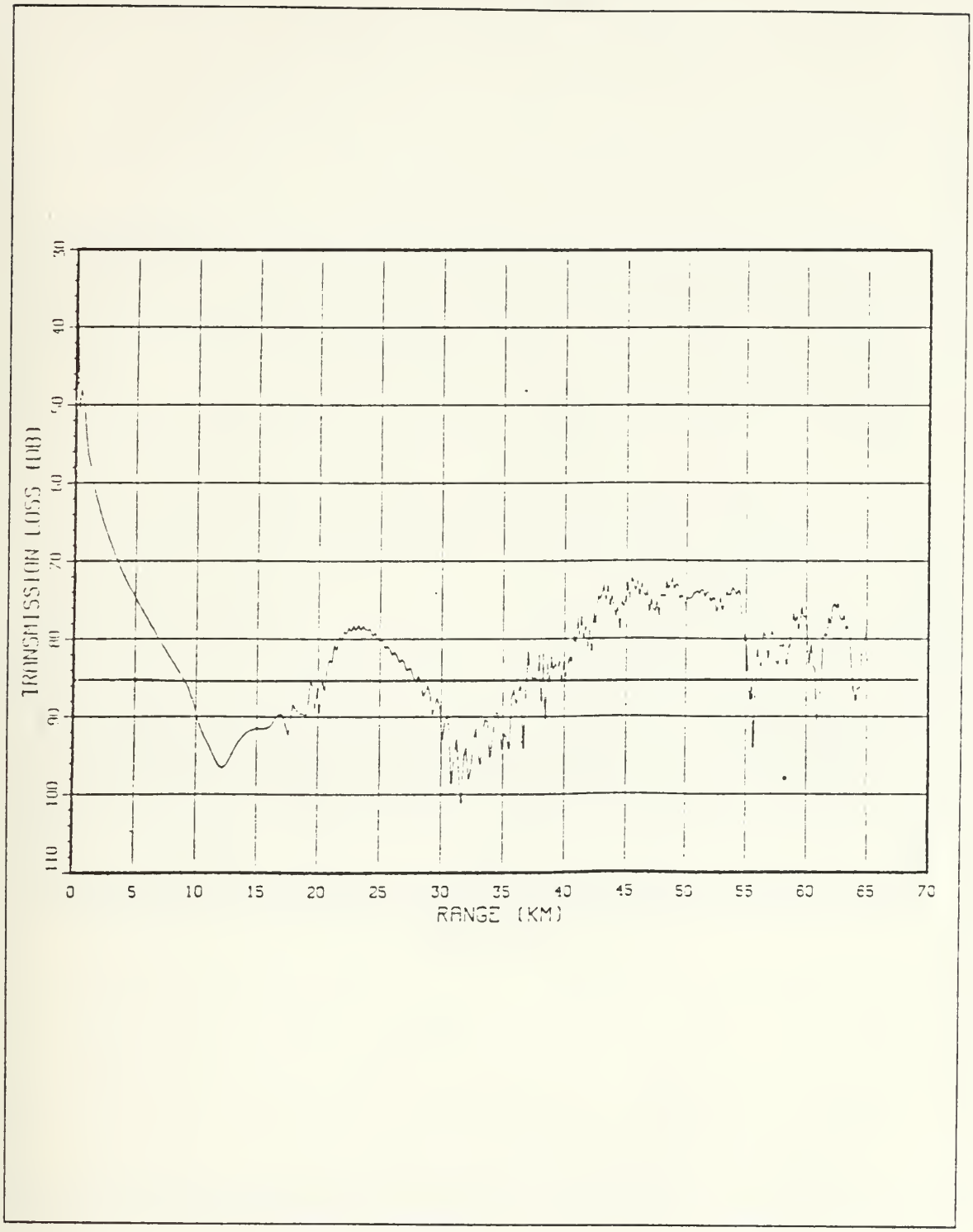


Figure 4.7 A PL curve for PE at 50 Hz using multiple sound speed profiles from a perpendicular transect (east to west). The source is at 60 m, the receiver is at 60 m.

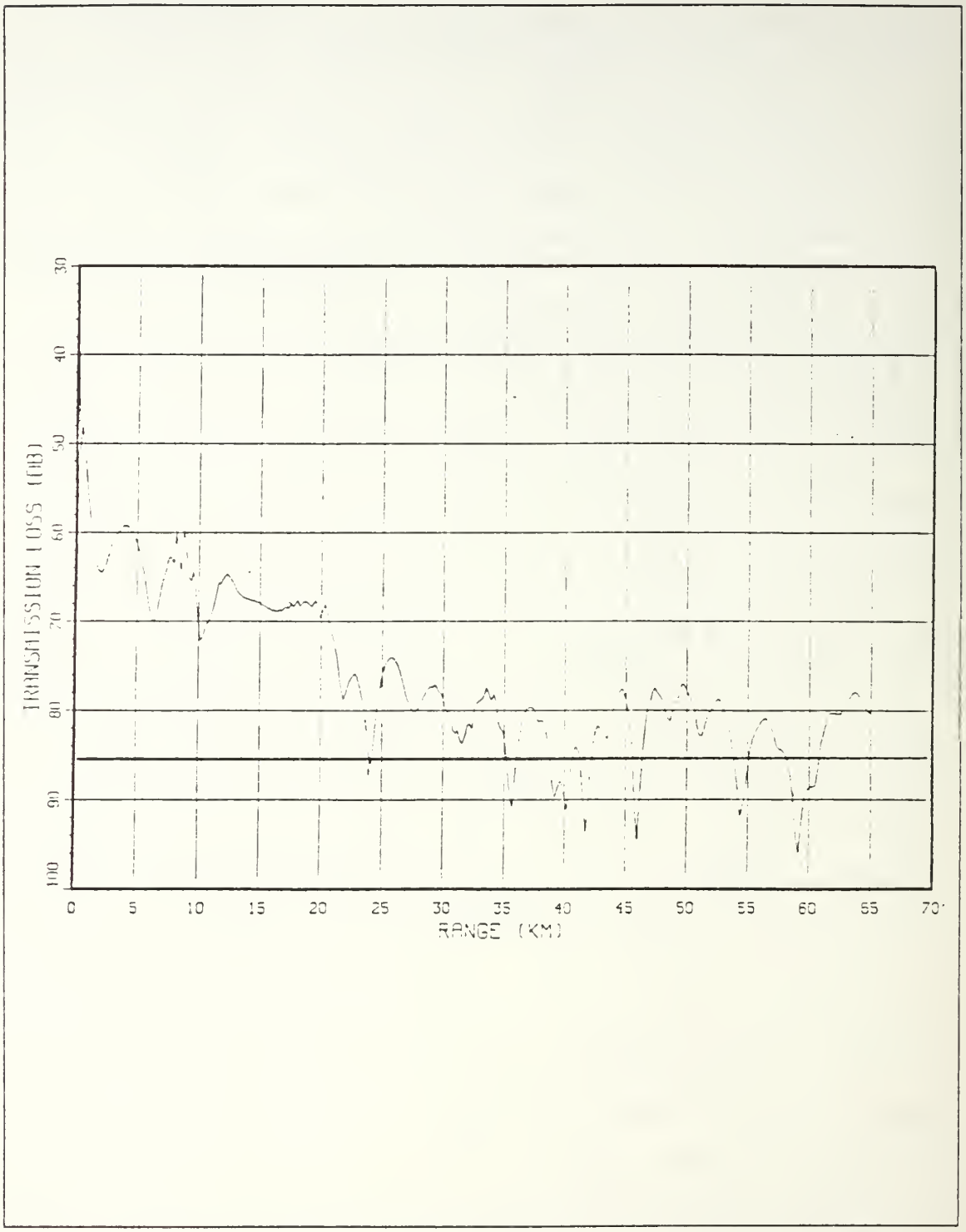


Figure 4.8 A PL curve for PE at 50 Hz using multiple sound speed profiles from a perpendicular transect (west to east). The source is at 60 m, the receiver is at 60 m.

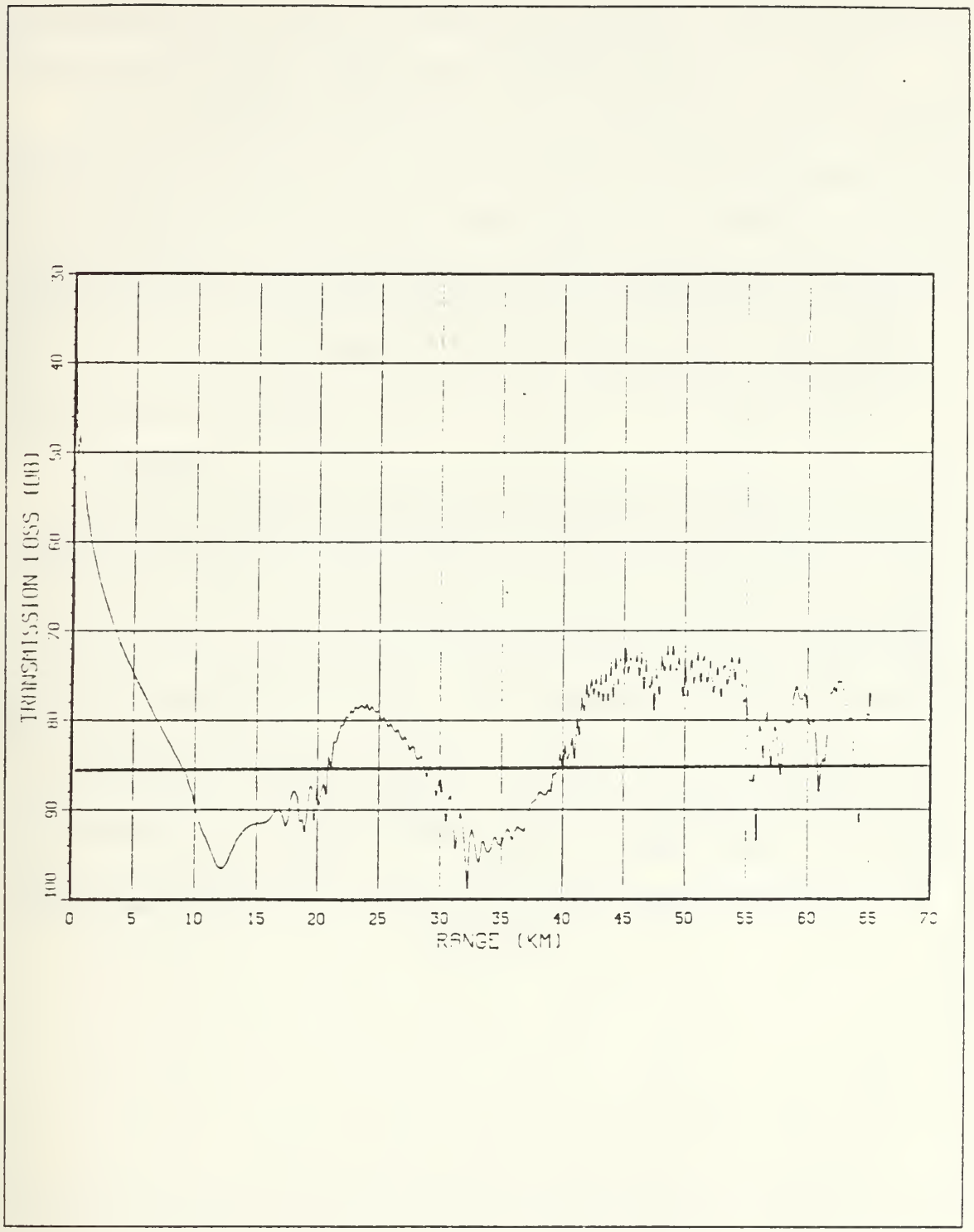


Figure 4.9 A PL curve for PE at 50 Hz using multiple sound speed profiles from a perpendicular transect (west to east), assuming a constant water depth. The source is at 60 m, the receiver is at 60 m.

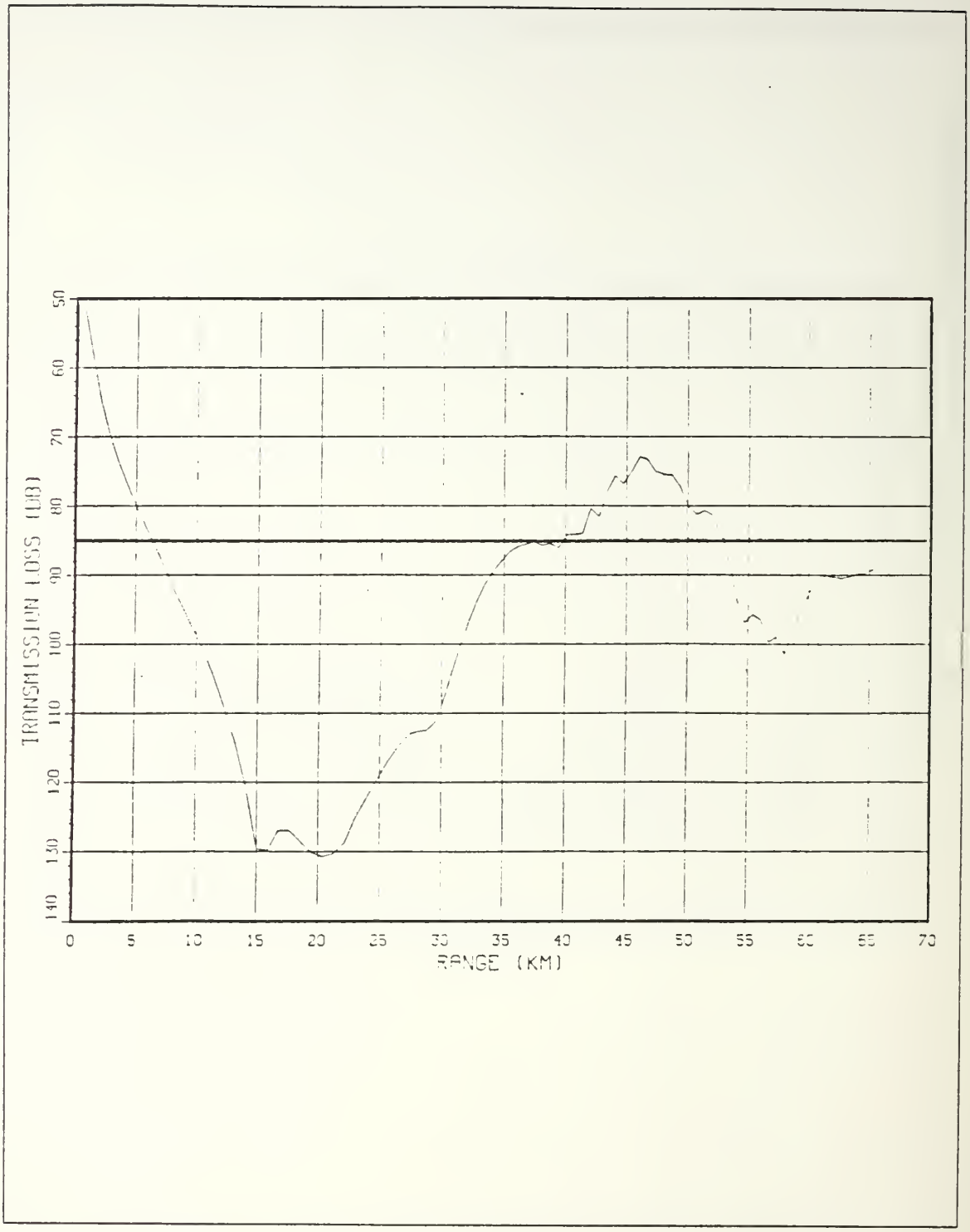


Figure 4.10 A PL curve for PE at 50 Hz using a single sound speed profile, assuming a sloping ocean bottom. The source is at 60 m, the receiver is at 60 m.

bottom effects come into play. The shoaling of the ocean floor produces the recovery of energy shown around 40 - 50 km, rising to a maximum signal excess of 10 dB.

Due to the significant differences in SSPs on either side of the EGPF and the strong influence of the source profile vis-a-vis the receiver profile, acoustic reciprocity does not pertain in the waters of the East Greenland Current. A shallow water source (i.e. one positioned on the continental shelf) is likely to be detected by a receiver operating in the deep waters to the east of the EGPF sooner than the reverse case. For the reverse case direct path propagation will be considerably lower, although CZ detection is possible. For example, with a FOM of 85 dB, a shallow water, continental shelf source is likely to be detected beyond 65 km, whereas for a deep water source detection is likely to only 8 km.

4. **Conclusions**

Because of its ability to incorporate range-varying parameters such as SSPs and bottom slopes, the PE model is the most suitable of the models considered. This range-dependent model is better suited to a range-dependent environment such as the EGPF region. FACT is less than ideal but gives plausible results in the shallow water environment and forecasts ranges that are optimistic by a factor of two in deep water. RAYMODE, while similar to FACT in shallow water, gave unrealistic results in the deep water environment.

Differences were noted between propagation normal and oblique to the EGPF, apparently due to SSP differences. However, in order to determine whether there is any significant difference in propagation at various aspects to the EGPF, would probably require a specific and carefully designed acoustic experiment.

# V. SUMMARY AND CONCLUSIONS

Two cluster analysis techniques, one heuristic and one iterative have been employed to investigate oceanographic data from the Greenland Sea. In particular, the techniques examined the natural groupings of the water masses of the East Greenland Current. Cluster techniques are not limited to temperature and salinity , but can accommodate any number of properties. Cluster analysis was succesful in identifying the natural groups, i.e., the water masses of the East Greenland Current. The techniques applied to various subsets of the EGC data proved to be robust and generally reliable. Of the two techniques, the iterative technique proved to be more consistent with classical oceanographic analysis. The heuristic technique, in which each entity is considered only once, was less successful in identifying the locus of the EGPF than the iterative technique. The cluster technique was shown to be simple in its applications and revealed the 'shape' (spatial groupings) of the data.

In addition, the technique was demonstrated on single attribute (variable) subsets, in particular temperature data. The results showed that cluster analysis using single-attribute data has useful applications in providing a quick categorisation of an ocean area. This should prove useful in obtaining some insight into the spatial coherence of selected areas. Such results would be useful in planning sonobuoy patterns or in determining the validity of XBT information prior to an acoustic forecast.

The acoustical analysis showed that acoustic reciprocity does not hold in the waters of the EGC. Ranges from shallow to deep water were far in excess of those from deep to shallow water. Propagation across the EGPF was shown to be different for normal and oblique cases. Oblique ranges were of the order of 80% of the orthogonal ranges when using a shallow water SSP. For deep to shallow water propagation, oblique ranges were of the order of 60% of the perpendicular ranges. The common shallow SSP has a very strong positive gradient causing significant focusing of energy. This ducting is continued along the track and only small differences in receiver SSPs are required to produce significantly different ranges. For deep to shallow water propagation the perpendicular and oblique cases had different source SSPs. The orthogonal profile had a much stronger surface duct than the

oblique profile. Energy was more quickly dissipated in the oblique case leading to the further reduction of transmission loss. The three acoustic models, FACT, RAYMODE and PE, all gave similar and generally reliable results when using shallow water SSPs. However, when using deep water SSPs for east to west propagation across the EGPF, there were significant differences. RAYMODE was extremely optimistic in its range prediction and FACT gave ranges that were twice those of the assumed best model. The overall conclusion of the acoustical analysis was that in such a range-dependent environment as the EGC one needs a range-dependent acoustic program, such as PE.

## LIST OF REFERENCES


Aagaard, K., and L.K. Coachman, The East Greenland Current North of the Denmark Strait, I, *Arctic*, 21(3), 181-200, 1968a.

Aagaard, K., and L.K. Coachman, The East Greenland Current North of the Denmark Strait, II, *Arctic*, 21(4), 267-290, 1968b.

Aagaard, K., J.H. Swift, and E.C. Carmack, Thermohaline Circulation in the Arctic Mediterranean Seas, *J. Geophys. Res.* 90,(C3), 4833-4846, 1985.

Bourke, R.H., and R.G. Paquette, MIZLANT 84 Data Report, Tech. Rep. NPS 68-85-018, Naval Postgraduate School, Monterey, California, 1985.

Brady, B.J., A Critical Analysis of Ocean Thermal Models in operation at FNOC, Masters Thesis, Naval Postgraduate School, Monterey, California, September 1985.

Brock, H.K., The AESD Parabolic Equation Model NORDA Technical Note 12, Naval Ocean Research and Development Activity, Bay St. Louis, Mississipi, 1978.

Chatfield, C., and A.J. Collins, *Introduction to Multivariate Analysis*, Chapman and Hall, New York, pp. 246, 1980.

Lorr, M., *Cluster Analysis for Social Scientists*, Josey-Bass, San Francisco, pp. 233, 1983.

Marsh, H.W. and M. Schulkin, Report on the Status of Project AMOS, Naval Underwater Sound Laboratory, New London, Connecticut, 1967.

Monsaigneon, L., Ocean Thermal Analysis and Related Naval Operational Considerations in the Ionian Sea, June 1980, Masters Thesis, Naval Postgraduate School, Monterey, California, September 1981.

Mooers, C.K., Notes on the Empirical Orthogonal Functions, Class notes for Synoptic Mesoscale Oceanography, Naval Postgraduate School, Monterey, California, 1985.

Paquette, R.G., R.H. Bourke, J.L. Newton, and W.F. Perdue, The East Greenland Polar Front in Autumn, *J. Geophys. Res.*. 90,(C3), 4866-4882, 1985.

Preisendorfer, R.W., Notes on Empirical Orthogonal Functions, Class notes for Time-Series Analysis, Naval Postgraduate School, Monterey, California, 1982.

RAYMODE, Passive Propagation Loss Program Performance Specification, *Document No. 8205*, New England Technical Services, Portsmouth, Rhode Island, 1982.

Sleichter, W.T., Modeling Acoustic Propagation Across the East Greenland Polar Front, Masters Thesis, Naval Postgraduate School, Monterey, California, December 1984.

Spath, H., *Cluster Analysis Algorithms*, Ellis Horwood Limited, Chichester, pp. 185, 1980.

Spofford, C.W., The FACT Model, Volume 1, Acoustic Environmental Support Detachment, Washington, DC., 1974.

Spofford, C.W., Interim Bottom-Loss Values and Geographic Assignment Based Upon Sediment Thickness, Science Applications Inc., Mclean, VA., 1980.

Swift, J.H., K. Aagard and S-A. Malmberg, The Contribution of the Denmark Strait Overflow to the Deep North Atlantic, *Deep-Sea Research*, 27A, 29-42, 1980.

Tunnicliffe, M.D., An Investigation into the Waters of the East Greenland Current Masters Thesis, Naval Postgraduate School, Monterey, California, September 1985.

UNESCO 1981, Tenth Report of the Joint Panel in Oceanographic Tables and Standards, *Unesco Tech. Pap. in Mar. Sci*, No. 36, 15-18, 1981.

Urick, R.J., *Principles of Underwater Sound*, McGraw-Hill, New York, pp. 423, 1983.

INITIAL DISTRIBUTION LIST

| | No. Copies |
|---|---|
| 1. Defense Technical Information Center<br>Cameron Station<br>Alexandria, VA 22304-6145 | 2 |
| 2. Library, Code 0142<br>Naval Postgraduate School<br>Monterey, CA 93943-5002 | 2 |
| 3. Chairman, Code 68Tm<br>Department of Oceanography<br>Naval Postgraduate School<br>Monterey, CA 93943 | 1 |
| 4. Chairman, Code 63Rd<br>Department of Meteorology<br>Naval Postgraduate School<br>Monterey, CA 93943 | 1 |
| 5. Dr. R H Bourke, Code 68Bf<br>Department of Oceanography<br>Naval Postgraduate School<br>Monterey, CA 93943 | 5 |
| 6. Dr. L C Breaker, Code 68By<br>Department of Oceanography<br>Naval Postgraduate School<br>Monterey, CA 93943 | 1 |
| 7. Director of Naval Oceanography and Meteorology<br>Lacon House<br>Theobalds Road<br>London WC1X 8RY<br>England | 4 |
| 8. Officer-in-Charge<br>Oceanographic Centre<br>CINCFLEET<br>Northwood<br>Middlesex<br>England | 2 |
| 9. Lt Cdr J M Clipson Royal Navy<br>Oceanographic Centre<br>CINCFLEET<br>Northwood<br>Middlesex<br>England | 1 |
| 10. Director Naval Oceanography Division<br>Naval Observatory<br>34th and Massachusetts Avenue NW<br>Washington, DC 20390 | 1 |
| 11. Commander<br>Naval Oceanography Command<br>NSTL Station<br>Bay St. Louis, MS 39522 | 1 |

12. Commanding Officer
Naval Oceanographic Office
NSTL Station
Bay St. Louis, MS 39522 — 1

13. Commanding Officer
Fleet Numerical Oceanography Center
Monterey, CA 93943 — 1

14. Commanding Officer
Naval Ocean Research and Development Activity
NSTL Station
Bay St. Louis, MS 39522 — 1

15. Commanding Officer
Naval Environmental Prediction
Research Facility
Monterey, CA 93943 — 1

16. Chairman, Oceanography Department
U. S. Naval Academy
Annapolis, MD 21402 — 1

17. Chief of Naval Research
800 N. Quincy Street
Arlington, VA 22217 — 1

18. Office of Naval Research (Code 420)
800 N. Quincy Street
Arlington, VA 22217 — 2

19. Scientific Liaison Office
Office of Naval Research
Scripps Institution of Oceanography
La Jolla, CA 92037 — 1

20. Library
Scripps Institution of Oceanography
PO Box 2367
La Jolla, CA 92037 — 1

21. Library
CICESE
PO Box Box 4803
San Ysidro, CA 92073 — 1

22. Library
School of Oceanography
Oregon State University
Corvallis, OR 97331 — 1

23. Commander
Oceanographic Systems Pacific
Box 1390
Pearl Harbor, HI 96860 — 1

24. Commander (AIR-370)
Naval Air Systems Command
Washington, DC 20360 — 1

25. Commanding Officer
Naval Eastern Oceanography Center
Naval Air Station
Norfolk, VA 23511 — 1

26. Commanding Officer
Naval Western Oceanography Center
Box 113
Pearl Harbor, HI 96860 — 1

27. Commanding Officer 1
Naval Oceanography Command Center, Rota
Box 31
FPO San Francisco, CA 09540

28. Commanding Officer 1
Naval Oceanography Command Center, Guam
Box 12
FPO San Francisco, CA 96630

29. Director 1
Arctic Submarine Laboratory
Code 54, Building 371
Naval Ocean Systems Center
San Diego, CA 92152

DUDLEY KNOX LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA 93943-5002

220146

Thesis
C516 Clipson
c.1 Water mass and acoustic analysis of the East Greenland Current.

220146

Thesis
C516 Clipson
c.1 Water mass and acoustic analysis of the East Greenland Current.

thesC516
Water mass and acoustic analysis of the
3 2768 000 68134 0
DUDLEY KNOX LIBRARY